

COMPARACIÓN DE LAS PRUEBAS DE PERMUTACIÓN Y CLÁSICAS  
(ASINTÓTICAS) APLICADAS EN TABLAS DE CONTINGENCIA DE  
DIMENSIÓN  $R \times C$  SIMPLE Y DOBLEMENTE ORDENADAS

Ing. Agr. Luis M. Pérez Y.  
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 LAS PRUEBAS DE PERMUTACIÓN Y CLÁSICAS  
(ASINTÓTICAS) APLICADAS EN TABLAS DE CONTINGENCIA DE  
DIMENSIÓN  $R \times C$  SIMPLE Y DOBLEMENTE ORDENADAS

Ing. Agr. Luis M. Pérez Y.

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**

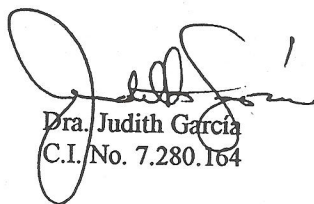
Quienes suscriben, Miembros del Jurado designado por el Consejo de la Facultad de Agronomía de la Universidad Central de Venezuela, para examinar el Trabajo de Grado presentado por el ciudadano **LUIS MANUEL PÉREZ YBARRA**, C.I. No. 12.195.023, titulado “**COMPARACIÓN DE LAS PRUEBAS DE PERMUTACIÓN Y CLÁSICAS (ASINTÓTICAS) APLICADAS EN TABLAS DE CONTINGENCIA DE DIMENSIÓN  $R \times C$  SIMPLE Y DOBLEMENTE 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 lunes 24 de octubre de 2005, a las 08:00 a.m., para que el autor lo defendiera en forma pública, lo que éste 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 el autor, que se ajusta a lo dispuesto y exigido por el Reglamento de Estudios de Postgrado. Para dar este Veredicto el jurado estimó que el trabajo examinado constituye un aporte en el análisis de datos categorizados ya que presenta un estudio comparativo entre pruebas de permutación y asintóticas aplicadas a tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenada, lo cual por un lado permitió cumplir con los objetivos propuestos y por el otro representa una importante referencia en el área de la estadística aplicada. El Jurado quiere dejar constancia de la calidad de la exposición y defensa pública que realizó el autor.

Página 2 del Veredicto del Trabajo de Grado del ciudadano LUIS MANUEL PÉREZ YBARRA.

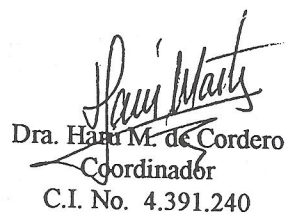
En fe de lo cual se levanta la presente acta, a los veinticuatro días del mes de octubre de dos mil cinco, dejándose también constancia de que, conforme a lo dispuesto en la normativa jurídica vigente actuó como Coordinador del Jurado la tutora, Dra. Harú Martínez de Cordero.

Firma del Jurado Examinador:

  
Dra. Judith García  
C.I. No. 7.280.164



  
Dr. Lusbi Herrera  
C.I. No. 3.255.060

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

24-10-05  
Veredicto

Miriam

*“By the turn of the century, we should see advances in applicability of exact methodology for contingency tables at least comparable to those of the past decade. One does not need a crystal ball to predict that computer speed will continue to increase and algorithms will be further improved, so that tables not now feasible for analysis soon will be”.*

Alan Agresti. 1.992.

## **DEDICATORIA**

A mis padres Luis Enrique y Carmen Alicia.

A mi hermano Carlos Enrique.

A mis hermanas Marianela, Grizly y Génesis.

## **AGRADECIMIENTOS**

A mis padres, por su incondicional apoyo durante toda mi formación de pre y postgrado y especialmente por su paciencia durante la elaboración de este trabajo de grado.

A mi tutora, Dra. Harú Martínez de Cordero, por estar siempre presente como tutora y como amiga, siempre apoyándome, motivándome y dándome sabios y útiles consejos.

A Lourdes América, sin cuyo apoyo incondicional, amistad y responsabilidad en el trabajo hubiera sido imposible la consecución de esta importante meta en mi vida.

A los profesores miembros tanto del comité consejero como del jurado examinador por enriquecer este trabajo con sus observaciones y correcciones.

## 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	20
5.1. Antecedentes	20
5.1.1. Historia y trabajos previos con pruebas de permutación	20
5.1.2. Trabajos previos en datos categorizados	23
5.1.2.1. Alan Agresti y colaboradores	23
5.1.2.2. Cyrus Mehta, Nitin Patel y colaboradores	24
5.1.2.3. Otros aportes importantes	25



5.2. Generalidades sobre pruebas de permutación	26
5.3. Aspectos generales de las tablas de contingencia	30
5.3.1. Independencia de variables categóricas	33
5.3.2. Esquemas de muestreo para tablas de contingencia	34
5.3.2.1. Muestreo Poisson	34
5.3.2.2. Muestreo multinomial	34
5.1.2.3. Muestro de producto de multinomiales	35
5.4. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$	36
5.4.1. Medida de discrepancia	39
5.4.2. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$ simplemente ordenadas	39
5.4.2.1. Prueba de Kruskal-Wallis	42
5.4.2.2. Prueba de marcas normales	43
5.4.2.3. Prueba de marcas de Savage	44
5.4.2.4. Análisis de varianza de una vía con marcas arbitrarias	44
5.4.3. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$ doblemente ordenadas	48
5.4.3.1. Prueba de Jonckheere-Terpstra	50
5.4.3.2. Prueba de asociación lineal-lineal	52

5.5. Aspectos computacionales sobre pruebas de permutación en tablas de contingencia	55
5.5.1. Enumeración completa	55
5.5.2. Simulación	60
6. RESULTADOS Y DISCUSIÓN	64
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	64
6.1.1. Tablas de contingencia de dimensión $R \times C$ simplemente ordenadas	64
6.1.2. Tablas de contingencia de dimensión $R \times C$ doblemente ordenadas	80
6.2. Estudio del comportamiento de las metodologías exactas y asintóticas a partir de la simulación de tablas de contingencia	87
6.2.1. Simulación en tablas de contingencia de dimensión $R \times C$ simplemente ordenadas	90
6.2.2. Simulación en tablas de contingencia de dimensión $R \times C$ doblemente ordenadas	98
6.2.3. Tamaños de muestra minimax	102
7. CONCLUSIONES Y RECOMENDACIONES	106

8. APÉNDICE	109
8.1. Deducción de la distribución de probabilidad exacta para pruebas de permutación condicionadas en tablas de contingencia de dimensión $R \times C$ según el esquema de muestreo de producto de poblaciones multinomiales independientes	109
9. REFERENCIAS BIBIOGRÁFICAS	114
10. ANEXOS	123

## ÍNDICE DE TABLAS

	Página
<i>Tabla 1. Estructuras de los vectores de parámetros multinomiales</i>	12
<i>Tabla 2. Configuraciones poblacionales de las tablas de contingencia</i>	13
<i>Tabla 3. Totales marginales de las filas de las tablas de contingencia simuladas</i>	19
<i>Tabla 4. Cuadro comparativo de los diferentes enfoques inferenciales para probar igualdad de medias de dos poblaciones</i>	30
<i>Tabla 5. Tabla de contingencia de dimensión <math>R \times C</math> para las variables categóricas <math>X</math> e <math>Y</math></i>	31
<i>Tabla 6. Parámetros de la distribución de probabilidad conjunta bivariada de dimensión <math>R \times C</math> de las variables categóricas <math>X</math> e <math>Y</math></i>	32
<i>Tabla 7. Estado de pacientes con esclerosis múltiple crónica y progresiva al cabo de 20 semanas de tratamiento</i>	65
<i>Tabla 8. Resultados de los <math>p</math>-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 1</i>	65
<i>Tabla 9. Pacientes dados de alta vivos o muertos, de acuerdo a las unidades de volumen de émbolos presentados en aneurismas aórtico</i>	67
<i>Tabla 10. Resultados de los <math>p</math>-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 2</i>	68
<i>Tabla 11. Calidad de la visión del ojo derecho en hombres y mujeres</i>	69
<i>Tabla 12. Resultados de los <math>p</math>-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 3</i>	70

<b>Tabla 13.</b> <i>Extensión del crecimiento del tumor pulmonar según el estilo de vida del paciente</i>	71
<b>Tabla 14.</b> <i>Resultados de los p-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 4</i>	72
<b>Tabla 15.</b> <i>Malformaciones congénitas en los órganos genitales de los recién nacidos según el consumo de alcohol de las madres durante los tres primeros meses de embarazo</i>	74
<b>Tabla 16.</b> <i>Resultados de los p-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 5</i>	75
<b>Tabla 17.</b> <i>Tiempo de duración del hábito de fumar y presencia de cáncer cervical</i>	78
<b>Tabla 18.</b> <i>Resultados de los p-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 6</i>	78
<b>Tabla 19.</b> <i>Tabla de contingencia doblemente ordenada utilizada por Agresti y Wackerly como ejemplo ilustrativo</i>	80
<b>Tabla 20.</b> <i>Resultados de los p-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 7</i>	80
<b>Tabla 21.</b> <i>Severidad de dumping y tipo de operación realizada</i>	83
<b>Tabla 22.</b> <i>Resultados de los p-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 8</i>	83
<b>Tabla 23.</b> <i>Calificación del docente y condiciones en la que se realizó determinada tarea</i>	85
<b>Tabla 24.</b> <i>Resultados de los p-valores exactos y asintóticos con el <math>\Delta</math> correspondiente para los datos del ejemplo 9</i>	85

<b>Tabla 25.</b> <i>Tamaños de muestra críticos observados para el umbral <math>\Delta^* = 0,01</math> en la prueba de Kruskal-Wallis para tablas de contingencia simplemente ordenadas</i>	94
<b>Tabla 26.</b> <i>Tamaños de muestra críticos observados para el umbral <math>\Delta^* = 0,01</math> en la prueba de marcas normales para tablas de contingencia simplemente ordenadas</i>	95
<b>Tabla 27.</b> <i>Tamaños de muestra críticos observados para el umbral <math>\Delta^* = 0,01</math> en la prueba de marcas de Savage para tablas de contingencia simplemente ordenadas</i>	96
<b>Tabla 28.</b> <i>Tamaños de muestra críticos observados para el umbral <math>\Delta^* = 0,01</math> en el análisis de varianza de una vía con marcas arbitrarias para tablas de contingencia simplemente ordenadas</i>	97
<b>Tabla 29.</b> <i>Tamaños de muestra críticos observados para el umbral <math>\Delta^* = 0,01</math> en la prueba de Jonckheere-Terpstra para tablas de contingencia doblemente ordenadas</i>	100
<b>Tabla 30.</b> <i>Tamaños de muestra críticos observados para el umbral <math>\Delta^* = 0,01</math> en la prueba de asociación lineal-lineal para tablas de contingencia doblemente ordenadas</i>	101
<b>Tabla 31.</b> <i>Tamaños de muestra minimax observados con el umbral <math>\Delta^* = 0,01</math> para las pruebas consideradas en tablas de contingencia de dimensión <math>R \times C</math> simplemente ordenadas</i>	104
<b>Tabla 32.</b> <i>Tamaños de muestra minimax observados con el umbral <math>\Delta^* = 0,01</math> para las pruebas consideradas en tablas de contingencia de dimensión <math>R \times C</math> doblemente ordenadas</i>	105

<b>Tabla 33.</b> <i>Tabla de contingencia de dimensión <math>R \times C</math></i>	109
<b>Tabla 34.</b> <i>Tabla de contingencia observada de dimensión <math>2 \times 2</math></i>	110

## ÍNDICE DE FIGURAS

	Página
<i>Figura 1. Valores de la medida de similitud <math>\Delta</math> según el tamaño de muestra para las pruebas consideradas en tablas de contingencia de dimensión <math>3 \times 3</math> simplemente ordenadas desbalanceadas y vectores de probabilidades multinomiales simétricos</i>	89



## RESUMEN

En este trabajo se realiza un estudio comparativo entre pruebas de permutación y asintóticas, aplicadas a tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas, utilizando como medida de comparación la diferencia entre el  $p$ -valor exacto y asintótico.

Se analizaron ejemplos, que usan el enfoque asintótico, publicados en la literatura científica internacional, con el objeto de conocer las condiciones bajo las cuales los  $p$ -valores asintóticos asociados a las pruebas consideradas: prueba de Kruskal-Wallis, marcas normales, marcas de Savage y ANAVAR con marcas arbitrarias; convergen a sus contrapartes exactos obtenidos por pruebas de permutación. Adicionalmente, se simularon tablas de contingencia, bajo el esquema de muestreo multinomial, para determinar el tamaño de muestra a partir del cual los resultados obtenidos por ambos métodos son equivalentes; es decir, cuando puede reportarse el  $p$ -valor asintótico en vez del  $p$ -valor exacto.

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, elección de las marcas en filas y columnas y balanceo de los datos; y que el tamaño de muestra a partir del cual ambas metodologías son equivalentes dependió de la prueba utilizada y de la dimensión de la tabla de contingencia considerada, variando desde tamaños de muestra tan pequeños como  $N = 40$  hasta tan grandes como  $N > 1000$ .

**Palabras claves:** pruebas de permutación, pruebas asintóticas, tablas de contingencia simplemente ordenadas, tablas de contingencia doblemente ordenadas,  $p$ -valor, marcas de filas y columnas, esquema de muestreo multinomial.

## ABSTRACT

In this work is carried out a comparative study between permutation and asymptotic tests applied to  $R \times C$  dimension singly and doubly ordered contingency tables, by using the difference between the exact  $p$ -value and asymptotic one as a comparison measure.

Examples published in the international scientific literature that use the asymptotic approach were analyzed in order to knowing the conditions under those which the asymptotic  $p$ -values associated to the considered tests: Kruskal-Wallis test, normal scores, Savage scores and ANOVA with arbitrary scores; converge to their counterpart exacts obtained by permutation tests. Additionally, contingency tables were simulated, under the scheme of multinomial sampling, to determine the sample size starting from which the results obtained by both methods are equivalent; that is, when the asymptotic  $p$ -value can be reported instead of the exact  $p$ -value.

The results showed that the differences among the  $p$ -values associated to the hypothesis tests applied ranked according to the contingency table dimension, sample size, election of the row and column scores and data balance; and the sample size starting from which both methodologies are equivalent depended on the used test and of the dimension of the contingency table considered, ranking from sample sizes so small as  $N = 40$  until as large as  $N > 1000$ .

***Key words:* permutation tests, asymptotic tests, simply ordered contingency tables, doubly ordered contingency tables,  $p$ -value, row and column scores, multinomial sampling.**

## **1. INTRODUCCIÓN**

### **1.1. Planteamiento del problema**

En muchos trabajos de investigación en los cuales tanto la variable causa como la respuesta son de naturaleza discreta y hasta cualitativa, resulta conveniente presentar los datos agrupados en tablas de contingencia, específicamente cuando la respuesta medida esté constituida por frecuencias absolutas y se desee relacionar a las variables causa y respuesta; por ejemplo, si se desea medir la opinión de diferentes sectores del país nacional con respecto a alguna política emanada por el ejecutivo, se mide la frecuencia de las opiniones de los diferentes sectores y se agrupan en una o más tablas de contingencia.

Por otra parte, muchas veces ocurre que a pesar de ser la variable respuesta cuantitativa e incluso continua, la naturaleza de la investigación impide que los datos originales se tomen directamente; en vez de ello, se considera la frecuencia absoluta de éstos en unas clases preestablecidas; por ejemplo, en un estudio socioeconómico; por lo general, no se le pregunta el sueldo exacto al encuestado, sino que se le propone elegir entre diferentes opciones salariales presentadas en clases, y luego a partir de tal distribución de frecuencias se elaboran las tablas de contingencia respectivas.

Si bien existen métodos para analizar tales resultados, por lo general éstos se basan en algún estadístico cuya distribución asintótica es conocida y se encuentra

tabulada; por ejemplo, la muy conocida prueba de independencia basada en el estadístico de Pearson y su correspondiente distribución  $\chi^2$ ; tales métodos se comportan adecuadamente en presencia de muestras grandes y balanceadas; sin embargo, esto no necesariamente ocurre si el tamaño de muestra es pequeño o si es grande pero con un conjunto de datos desbalanceado y/o disperso, (Agresti; 2002, Cytel Software, 2001 y 2003; Good, 2000). Por otra parte, muchos de los paquetes informáticos estadísticos que se utilizan frecuentemente, tales como SPSS®, Statistix®, Minitab®, SAS®, Statistica®, Systat®, etc.; no ejecutan el análisis correspondiente si no están presentes un mínimo de frecuencias por celda o advierten que tal aproximación es inválida.

En el caso de muestras pequeñas, Agresti y Wackerly, (1977); y Baglivo, Olivier y Pagano, (1988); indican que cuando no se cumplen los supuestos para aplicar el enfoque asintótico, muchas veces el investigador ignora, elimina o agrupa de forma conveniente a las clases de las tablas de contingencia para obtener una nueva tabla que pueda ser analizada con métodos asintóticos; el problema que se presenta con esta manipulación del conjunto original de datos, es que la tabla resultante no necesariamente se comporta de forma acorde con la tabla original, e incluso, muchas veces este comportamiento resulta diferente, lo cual frecuentemente conduce al investigador a formular conclusiones erróneas. Una alternativa para dar solución a estos problemas lo constituyen las pruebas de

permutación aplicadas a datos categorizados, las cuales generalmente utilizan el mismo estadístico que las correspondientes aproximaciones asintóticas pero que a diferencia de éstas, utilizan la distribución de las permutaciones de los datos en vez de la distribución teórica correspondiente; lo cual permite que la prueba realizada a partir del enfoque basado en pruebas de permutación presente niveles de significación exactos, dado que éstos se calculan utilizando una distribución discreta y libre de parámetros desconocidos generada a partir de los datos originales.

Las pruebas de permutación exigen alta capacidad de procesamiento computacional, lo cual constituye una limitación práctica de las mismas en contraposición con el enfoque clásico en el cual la comparación se establece con una distribución teórica donde su exigencia computacional es mínima, (Good, 2000); sin embargo, las condiciones económicas o la naturaleza de la población impiden muchas veces que se pueda tomar una muestra grande; en estos casos, la opción más adecuada para realizar inferencias confiables lo constituyen las pruebas de permutación, a pesar de los exigentes procesos computacionales que estas pruebas precisan.

En el presente trabajo se compara el comportamiento de los diferentes enfoques, clásico y exacto, en pruebas aplicadas a tablas de contingencia simple y doblemente ordenadas bajo diferentes escenarios, a saber: muestras pequeñas,

muestras grandes pero desbalanceadas y dispersas y muestras grandes y balanceadas; con la intención de mostrar en cuales escenarios es preferible un enfoque u otro.

## **1.2. Justificación de la investigación**

Puesto que en muchos estudios los datos se agrupan en tablas de contingencia, es evidente que un análisis adecuado de éstas proporcionará información útil para el análisis de tales variables; sin embargo, la literatura muestra que los métodos clásicos de amplia utilización en estos estudios constituyen solamente aproximaciones basadas en la ley de los grandes números a distribuciones teóricas tabuladas. Good, (2000); señala que si la muestra es pequeña o irregularmente distribuida tales aproximaciones son por lo general inexactas y en consecuencia, poco útiles para realizar inferencias confiables a partir de tal muestra; en cambio, utilizando pruebas de permutación se logra trabajar con un nivel de significación exacto, independientemente de la forma de la distribución y/o del tamaño de la muestra.

Por otra parte, Bishop, Fienberg y Holland, (1975); citados por Good, (2000); señalan que a medida que se incrementa el tamaño de la muestra las pruebas de permutación convergen a la distribución tabulada y así, en muestras grandes ambos enfoques son equivalentes y por lo tanto el uso de uno u otro enfoque depende del investigador.

Así pues, un estudio comparativo de ambos enfoques bajo los escenarios mencionados anteriormente constituye una forma de mostrar el comportamiento disímil reportado por la literatura para tales métodos, además de permitir dar recomendaciones acerca de cuando es preferible el uso de uno u otro.

### **1.3. Alcance**

El estudio abarca las metodologías para realizar inferencias estadísticas exactas en datos categorizados agrupados en tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas, con la intención de dar a conocer una alternativa que es aplicable en aquellas situaciones en las cuales los métodos tradicionales fallan y por ende, no producen resultados confiables.

En la actualidad, existen una serie de paquetes informáticos estadísticos que llevan a cabo pruebas de hipótesis con métodos exactos tanto para datos categorizados, como para pruebas de rangos y análisis de varianza, entre ellos se pueden citar:

- StatXact®.
- LogXact®.
- SPSS Exact®.
- S-Plus®.
- StatXact para SAS®.

- LogXact para SAS®.
- PermutIt®.

#### **1.4. Limitaciones**

Una limitación para la aplicación de los métodos exactos, lo constituye el hecho de que dependen del poder de procesamiento que posea el computador utilizado, de manera tal que el rango de aplicación de estas metodologías variará de un usuario a otro; aquellos usuarios que dispongan de computadoras más potentes podrán analizar un mayor número de situaciones que aquellos poseedores de computadoras menos potentes; de hecho, Good, (2000); comenta que el desarrollo de estas metodologías va paralelo a la evolución tanto del poder de procesamiento del *hardware* como del *software* y los algoritmos creados para tal fin.



## 2. OBJETIVOS

### 2.1. Objetivo general

- Comparar las pruebas de permutación y clásicas (asintóticas) aplicadas en tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas.

### 2.2. Objetivos específicos

- Comparar desde el punto de vista teórico a 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  $R \times C$  simple y doblemente ordenadas.
- Determinar el  $p$ -valor obtenido por la metodología asintótica para tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas.
- Comparar los  $p$ -valores obtenidos por la metodología asintótica y la correspondiente a pruebas de permutación para tablas de contingencia de dimensión  $R \times C$  simple y doblemente 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 obtenidos a partir de pruebas de hipótesis aplicadas a tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas, utilizando pruebas de permutación y métodos asintóticos presentarán, en general, valores diferentes para los conjuntos de datos pequeños; grandes y desbalanceados; grandes y dispersos; y un comportamiento similar para los conjuntos de datos grandes y balanceados.

## 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 aplicadas a tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas.

Las pruebas que se van a aplicar en el caso de las tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas son:

- La prueba de Kruskal-Wallis.
- La prueba de marcas (*scores*) normales.
- La prueba de marcas de Savage.
- Análisis de varianza de una vía (ANAVAR) con marcas arbitrarias.

Y para el caso de las tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas:

- La prueba de Jonckheere-Terpstra.
- La prueba de asociación lineal-lineal.

Los ejemplos considerados se seleccionaron de tal forma que permitieran mostrar las condiciones bajo las cuales el comportamiento es disímil para ambos enfoques, para tal fin, 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 calcularon 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.

Para mostrar como la tasa de convergencia del enfoque asintótico al exacto se ve afectada por características intrínsecas a la tabla; es decir, dimensión, tamaño de muestra, vectores de parámetros multinomiales, configuración poblacional y balanceo; las tablas de contingencia simuladas fueron diseñadas considerando las siguientes estructuras:

- Las dimensiones consideradas en este estudio para las tablas de contingencia simplemente ordenadas variaron desde  $2 \times c$  hasta  $5 \times c$ ; con

$c = 3, 4, 5$ ; y para las tablas de contingencia doblemente ordenadas desde  $3 \times c$  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:  $N = 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500, 1000$ .
- 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 1.

*Tabla 1. Estructuras de los vectores de parámetros multinomiales.*

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

- A partir de estos vectores de probabilidad multinomial se crearon las configuraciones poblacionales para las tablas de contingencia según la hipótesis nula de igualdad entre los vectores de probabilidad multinomiales sea cierta o falsa, tales configuraciones se presentan en la tabla 2.

*Tabla 2. Configuraciones poblacionales de las tablas de contingencia.*

CONFIGURACIONES POBLACIONALES DE LAS TABLAS DE CONTINGENCIA				
Categorías de filas	H <sub>0</sub> Cierta			H <sub>0</sub> falsa
$r = 2$	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 multinomiales 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 3, mostrada al final de esta sección.

Todas estas posibles combinaciones de tablas de contingencia generadas hacen un total de 1404 tablas de contingencia simplemente ordenadas y 936 tablas de contingencia doblemente ordenadas.

Posteriormente, se generaron diez muestras aleatorias multinomiales independientes para cada uno de los totales marginales de filas y vectores de probabilidad multinomial 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 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, tal como lo establece Agresti, (2002).

Las instrucciones para generar las muestras de poblaciones multinomiales independientes utilizando el *software* Dataplot® se muestran en el anexo 1.

A cada una de estas tablas generadas por simulación se les aplicaron las pruebas asintóticas y exactas 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,  $\Delta$  tenderá a cero; y se espera que esto ocurra para tamaños de muestra grandes, es decir:

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

Berger, (2000); propone una estrategia basada en esta medida de similitud  $\Delta$  para decidir entre el uso de las pruebas 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  $\Delta^*$  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  $\Delta^*$ .

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  $\Delta$  observado en la simulación será menor o igual al  $\Delta^*$  establecido; y es partir de este tamaño de muestra 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  $b$  y prueba  $t$  y configuración de los vectores de parámetros multinomiales  $h$ .

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

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

para el caso de las tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas, y

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

para el caso de las tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas.

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

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

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

$$t = \left\{ \begin{array}{l} \text{Kruskal-Wallis, Marcas Normales} \\ \text{Marcas de Savage, Anova con marcas arbitrarias} \end{array} \right\}$$

para el caso de las tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas, y

$$t = \left\{ \begin{array}{l} \text{Jonckheere-Terpstra,} \\ \text{Asociación lineal-lineal} \end{array} \right\}$$

para el caso de las tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas.

$h$ : es la configuración poblacional de los vectores de parámetros multinomiales considerada, mostradas en la tabla 2.

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 para 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 multinomiales considerados.

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

*Tabla 3. Totales marginales de las filas de las tablas de contingencia simuladas.*

<b>TOTALES MARGINALES DE LAS FILAS DE LAS TABLAS DE CONTINGENCIA</b>								
<i>N</i>	$2 \times c$		$3 \times c$		$4 \times c$		$5 \times c$	
	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  $R \times C$  simple y doblemente ordenadas. En la sección 5.1 se hace una revisión bibliográfica de los trabajos previos en pruebas de permutación especialmente en su aplicación para datos categorizados; en la sección 5.2 se discuten aspectos generales sobre pruebas de permutación; en la sección 5.3 se describen las generalidades de las tablas de contingencia de dimensión  $R \times C$ ; en la sección 5.4 se estudian las pruebas de permutación aplicadas a tablas de contingencia de dimensión  $R \times C$  simple y doblemente 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  $2 \times 2$ .
- 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 estuvieron orientados 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 más destacados 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 con miras a su aplicación en 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 gran 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.
- Agresti, Mehta y Patel, (1990); utilizaron el algoritmo de red de Mehta y Patel y lo aplicaron a tablas de contingencia con categorías ordenadas.
- Agresti, Lang y Mehta, (1993); compararon el comportamiento de varias pruebas condicionadas en tablas ordenadas y concluyeron que la prueba de asociación lineal-lineal es conservadora, pero que este conservacionismo disminuye con la dimensión de la tabla y con el tamaño de la muestra.

#### 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 y Hilton, (1993); compararon la potencia de dos pruebas, una condicionada y otra no condicionada, para tablas  $2 \times 3$  y encontraron que la

potencia de ambas es similar e incluso superada por la prueba condicionada y indicaron 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.
- Cyrus Mehta es actualmente presidente de Cytel Software Corporation.
- Gajjar, Mehta, Patel y Senchaudhuri son los creadores del *software* StatXact®, (Cytel Software, 2001 y 2003).

#### 5.1.2.3. Otros aportes importantes

- 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.
- Graubard y Korn, (1987); propusieron criterios para la elección de las marcas adecuadas en tablas de contingencia ordenadas cuando se llevan a cabo pruebas de independencia.
- Baglivo, Oliver y Pagano, (1988); propusieron una metodología mixta, en parte exacta y en parte asintótica, para calcular  $p$ -valores aproximados en tablas de contingencia.

- Hollander y Wolfe, (1999); son autores del libro de texto "*Nonparametric Statistical Methods*", el cual incluye pruebas de permutación y constituye una de las contribuciones más sólidas para el estudio de la estadística no paramétrica, por presentar éste una amplia descripción de las pruebas no paramétricas desarrolladas hasta nuestros días.
- Berger, V. y colaboradores, trabajan principalmente en datos categorizados ordenados y recomiendan el uso de marcas adecuadas en tales análisis. (Permutt y Berger, 2000; Ivanova y Berger, 2001).

## **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 rearreglos.
- 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  $\alpha$ ; además, se dice que una prueba es conservadora si el error Tipo I nunca excede  $\alpha$ , (Good, 2000).

Se dice que una prueba es insesgada y de nivel  $\alpha$ , siempre que su función de potencia  $\beta_\theta$  para el parámetro  $\theta$  satisface las siguientes condiciones:

- (i)  $\beta_\theta$  es conservadora; es decir,  $\beta_\theta \leq \alpha$  para todo  $\theta$  que satisfaga la hipótesis nula.
- (ii)  $\beta_\theta \geq \alpha$  para todo  $\theta$  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 igual o menor que  $\alpha$ , lo cual no necesariamente ocurre con las aproximaciones asintóticas; sin embargo, no necesariamente constituyen las pruebas más potentes disponibles para una situación dada.

Good, (2000); indica que las pruebas no paramétricas de rangos son simplemente pruebas de permutación aplicadas a los rangos de las observaciones

en vez de a los valores originales. Si las observaciones son intercambiables, entonces las pruebas de rangos resultantes son exactas e insesgadas, aunque, generalmente, menos potentes que las pruebas de permutación equivalentes.

Para el caso de las pruebas paramétricas, Good (2000); señala que si la distribución del estadístico de prueba es correcta, entonces la prueba paramétrica por lo general es más potente que su correspondiente prueba de permutación; en caso contrario, la prueba paramétrica puede ser menos potente y es preferible la prueba de permutación.

Para 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).

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).

La tabla 4 esquematiza los pasos necesarios para la aplicación de pruebas de hipótesis según tres enfoques diferentes: permutación, rangos y paramétricas; lo

que permiten visualizar que las estructuras de tales pruebas, aunque similares, varían según el enfoque, (Good, 2000).

**Tabla 4.** Cuadro comparativo de los diferentes enfoques inferenciales para probar igualdad de medias de dos poblaciones.

<b>Permutación</b>	<b>Rangos (Por ejemplo, Wilcoxon)</b>	<b>Paramétrico (Por ejemplo, prueba de <math>t</math>)</b>
Escoger un estadístico de prueba	Escoger un estadístico de prueba	Escoger un estadístico de prueba cuya distribución pueda ser derivada analíticamente
(Por ejemplo, suma de las observaciones en la primera muestra)	(Por ejemplo, suma de rangos en la primera muestra)	(Por ejemplo, $t$ de Student)
Calcular el estadístico	Convertir a rangos Calcular el estadístico	Calcular el estadístico
¿Son las observaciones intercambiables?	¿Son las observaciones intercambiables?	¿Son las observaciones independientes? ¿Siguen la distribución especificada?
Derivar la distribución de las permutaciones de la muestra combinada	Utilizar tabla de la distribución de las permutaciones de los rangos	Utilizar la distribución tabulada
Comparar el estadístico con los percentiles de la distribución	Comparar el estadístico con los percentiles de la distribución	Comparar el estadístico con los percentiles de la distribución

Fuente: Good, 2000.

### 5.3. Aspectos generales de las tablas de contingencia

Sean  $X$  e  $Y$  dos variables categóricas,  $X$  con  $r$  categorías representadas en filas e  $Y$  con  $c$  categorías representadas en columnas. La clasificación de las observaciones sobre ambas categorías tienen  $R \times C$  combinaciones posibles. Una tabla rectangular que tenga  $r$  filas y  $c$  columnas para las categorías de  $X$  e  $Y$ , respectivamente, muestra la distribución de la respuesta  $(X, Y)$  de un elemento



elegido al azar. Cuando las respuestas corresponden a una muestra, a tal tabla se le denomina tabla de contingencia. Una tabla de contingencia con  $r$  filas y  $c$  columnas, se denomina tabla de contingencia de dimensión  $R \times C$  o simplemente tabla de contingencia  $R \times C$  y se ilustra en la tabla 5, (Agresti, 2002).

*Tabla 5. Tabla de contingencia de dimensión  $R \times C$  para las variables categóricas  $X$  e  $Y$ .*

<b>Respuesta</b>	<b>Columnas (Y)</b>				<b>Total</b>
<b>Filas (X)</b>	$y_{11}$	$y_{12}$	$\dots$	$y_{1c}$	$m_{1+}$
	$y_{21}$	$y_{22}$	$\dots$	$y_{2c}$	$m_{2+}$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	$y_{r1}$	$y_{r2}$	$\dots$	$y_{rc}$	$m_{r+}$
<b>Total</b>	$n_{+1}$	$n_{+2}$	$\dots$	$n_{+c}$	$N$

Fuente: Cytel Software, 2001 y 2003; Agresti, 2002.

Las frecuencias conjuntas  $y_{ij}$ ;  $i=1, 2, \dots, r$ ;  $j=1, 2, \dots, c$ ; de la tabla, representan las  $R \times C$  posibles respuestas que corresponden a las variables  $X$  e  $Y$  medidas sobre una muestra aleatoria de tamaño  $N$ .

Las totales marginales resultan de la suma de las frecuencias conjuntas, así

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

Donde las  $\{m_{i+}\}$ ;  $i=1, 2, \dots, r$ ; constituyen los totales marginales de las filas; las  $\{n_{+j}\}$ ;  $j=1, 2, \dots, c$ ; los totales marginales de las columnas de la tabla de

contingencia de dimensión  $R \times C$  y el subíndice “+” indica que la suma ha sido hecha sobre todos los niveles de la otra variable, (Agresti, 2002).

Sea  $\pi_{ij}$ , la probabilidad de que la respuesta  $(X, Y)$  ocurra en la fila  $i$  y la columna  $j$ . La distribución de probabilidad de los  $\{\pi_{ij}\}$  constituye la distribución conjunta de  $X$  e  $Y$ , como se muestra en la tabla 6.

**Tabla 6.** Parámetros de la distribución de probabilidad conjunta bivariada de dimensión  $R \times C$  de las variables categóricas  $X$  e  $Y$ .

Respuesta	Columnas (Y)				Total
Filas (X)	$\pi_{11}$	$\pi_{12}$	...	$\pi_{1c}$	$\pi_{1+}$
	$\pi_{21}$	$\pi_{22}$	...	$\pi_{2c}$	$\pi_{2+}$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	$\pi_{r1}$	$\pi_{r2}$	...	$\pi_{rc}$	$\pi_{r+}$
<b>Total</b>	$\pi_{+1}$	$\pi_{+2}$	...	$\pi_{+c}$	1

Fuente: Agresti, 2002.

Las probabilidades marginales resultan de la suma de las probabilidades conjuntas, de aquí se tiene

$$\pi_{i+} = \sum_{j=1}^c \pi_{ij}$$

$$\pi_{+j} = \sum_{i=1}^r \pi_{ij}$$

Donde las  $\{\pi_{i+}\}; i=1, 2, \dots, r$ ; constituyen las probabilidades marginales de las filas; las  $\{\pi_{+j}\}; j=1, 2, \dots, c$ ; las probabilidades marginales de las columnas de la tabla de distribución bivariada de dimensión  $R \times C$ , (Agresti, 2002).

### 5.3.1. Independencia de variables categóricas

Dos variables categóricas  $X$  e  $Y$  se dicen independientes si sus probabilidades conjuntas son igual al producto de sus probabilidades marginales, así,

$$\pi_{ij} = \pi_{i+}\pi_{+j}; \text{ para } i=1, 2, \dots, r; j=1, 2, \dots, c$$

La distribución de la probabilidad condicional de la  $j$ -ésima columna dada la  $i$ -ésima fila está dada por:

$$\pi_{j|i} = \pi_{ij}/\pi_{i+}; j=1, 2, \dots, c$$

Así, cuando  $X$  e  $Y$  son independientes, se cumple:

$$\pi_{j|i} = \pi_{ij}/\pi_{i+} = \pi_{i+}\pi_{+j}/\pi_{i+} = \pi_{+j}; \text{ para } i=1, 2, \dots, r$$

Asumiendo que  $X$  es la variable explicativa o causa e  $Y$  la variable respuesta, se observa que bajo independencia  $\{\pi_{j|1} = \pi_{j|2} = \dots = \pi_{j|r} = \pi_{+j}\}$ ; para  $j=1, 2, \dots, c$ ; es decir, la probabilidad de respuesta de cualquier columna dada es la misma en cada fila. Bajo este punto de vista, la independencia no es más que homogeneidad de las distribuciones condicionales, (Agresti, 2002).

### 5.3.2. Esquemas de muestreo para tablas de contingencia

Agresti, (2000); plantea que existen tres esquemas de muestreo que pueden representarse adecuadamente mediante el uso de tablas de contingencia: el muestreo Poisson, el muestreo multinomial y el muestreo de producto de multinomiales.

#### 5.3.2.1. Muestreo Poisson

Considera a las respuestas  $Y_{ij}$  como variables aleatorias de Poisson independientes con parámetros  $\mu_{ij}$ . La función de probabilidad conjunta para las respuestas  $y_{ij}$  viene dada por el producto de  $R \times C$  funciones de probabilidad de Poisson, así se obtiene

$$\Pr(Y = y_{ij}) = \prod_{i=1}^r \prod_{j=1}^c \exp(-\mu_{ij}) \mu_{ij}^{y_{ij}} / y_{ij}!$$

#### 5.3.2.2. Muestreo multinomial

Ocurre cuando el total de la muestra  $N$  es fijo, pero no lo son los totales marginales de filas y columnas; así los posibles totales de filas y columnas  $y_{ij}$  poseen distribución multinomial, en este caso

$$\Pr(Y = y_{ij}) = \frac{N!}{y_{11}! \cdots y_{rc}!} \prod_{i=1}^r \prod_{j=1}^c \pi_{ij}^{y_{ij}}$$

### 5.3.2.3. Muestreo de producto de multinomiales

Cuando la respuesta ocurre separadamente para cada categoría  $i$  en las filas de la variable explicativa  $X$  los totales marginales de las filas son fijos, pero no así los totales de las columnas. Suponiendo que las  $m_{i+}$  observaciones en  $Y$  al nivel  $i$  de  $X$  son independientes, cada una con distribución de probabilidad condicionada

$\{\pi_{j|1}, \pi_{j|2}, \dots, \pi_{j|r}\}$ . Las frecuencias  $\{y_{ij}, j = 1, 2, \dots, c\}$  que satisfacen  $\sum_{j=1}^c y_{ij} = m_{i+}$

tienen distribución multinomial de la forma

$$\Pr(Y = y_{ij}) = \frac{m_{i+}!}{\prod_{j=1}^c y_{ij}!} \prod_{j=1}^c \pi_{j|i}^{y_{ij}} ; j = 1, 2, \dots, c \quad (1)$$

para cada nivel  $i$  de la variable  $X$

Cuando las muestras para diferentes niveles de  $X$  son independientes, la función masal de probabilidad conjunta para el conjunto total de datos es el producto de distribuciones multinomiales dadas por la expresión (1) para cada nivel  $i$  de  $X$ , este esquema de muestreo es llamado muestreo multinomial independiente o muestreo producto de multinomiales, cuya distribución de probabilidad viene dada por la expresión

$$\Pr(Y = y_{ij}) = \prod_{i=1}^r \left( \frac{m_{i+}!}{\prod_{j=1}^c y_{ij}!} \prod_{j=1}^c \pi_{ji}^{y_{ij}} \right) = \frac{\left( \prod_{i=1}^r m_{i+}! \right) \left( \prod_{i=1}^r \prod_{j=1}^c \pi_{ji}^{y_{ij}} \right)}{\prod_{i=1}^r \prod_{j=1}^c y_{ij}!} = \prod_{i=1}^r m_{i+}! \prod_{i=1}^r \prod_{j=1}^c \frac{\pi_{ji}^{y_{ij}}}{y_{ij}!}$$

#### 5.4. Pruebas de permutación aplicadas a tablas de contingencia de dimensión

$R \times C$

Para la aplicación de las pruebas de permutación a tablas de contingencia de dimensión  $R \times C$  es importante conocer cuál es la naturaleza de las tablas que se van a analizar, puesto que de ésta depende cuál es la prueba más conveniente a utilizar. En este sentido, las tablas de contingencia  $R \times C$  pueden clasificarse en tres tipos:

- i. Tablas de contingencia no ordenadas.
- ii. Tablas de contingencia simplemente ordenadas.
- iii. Tablas de contingencia doblemente ordenadas.

Las tablas de contingencia no ordenadas son aquellas en las cuales tanto la variable causa  $X$  representadas por las  $r$  filas y la variable respuesta  $Y$  representada por las  $c$  columnas, están medidas en escala nominal; un ejemplo de tales tablas lo constituye la tabla de contingencia obtenida de relacionar la zona geográfica de residencia (central, centro occidental, oriental, sur) con la presencia de dengue (ninguna, clásico, hemorrágico) en Venezuela.

Las tablas simplemente ordenadas son aquellas en las cuales la variable causa  $X$  está medida en escala nominal y la variable respuesta  $Y$  está medida en escala al menos ordinal; un ejemplo lo constituye la tabla de contingencia generada por el estudio de diferentes tratamientos de quimioterapia, llámense  $T_1, T_2, T_3, T_4$  como variable  $X$  y el tiempo de remisión de cierto tipo de tumor (0-2 meses, 2-4 meses, 4-6 meses, sin remisión del tumor) como variable respuesta  $Y$ .

Las tablas doblemente ordenadas son aquellas en las cuales tanto la variable causa  $X$  como la variable respuesta  $Y$  están medidas en escalas al menos ordinales; un ejemplo lo constituye la tabla de contingencia generada por el estudio de diferentes dosis de droga, (100 mg., 200 mg., 300 mg., 400 mg.) como variable  $X$  y el grado de toxicidad en ratas de laboratorio (leve, moderado, severo, muerte) como variable respuesta  $Y$ , (Cytel Software, 2001 y 2003; Good 2000).

Para llevar a cabo pruebas de permutación en tablas de dimensión  $R \times C$  simple y doblemente ordenadas, se utilizan pruebas condicionadas, siguiendo el razonamiento de Mehta y Hilton, (1993); los cuales encuentran que las potencias de las pruebas condicionadas en tablas de contingencia de dimensión  $2 \times 3$  es similar e incluso supera a la potencia de las pruebas no condicionadas, ellos conjeturan que en la medida en que la dimensión de la tabla aumente, será más marcada esta diferencia favoreciendo a las pruebas condicionadas.

El conjunto de referencia  $\Gamma$  es aquel formado por todas las tablas de contingencia  $\mathbf{y}_{r \times c}$  sujetas a la condición de que los totales de filas y columnas sean iguales a los de la tabla de contingencia observada, es decir,

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

donde  $m_{i+}$  y  $n_{+j}$  son los totales marginales de fila y columna, respectivamente, de la tabla de contingencia observada.

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; Agresti, 2002).

Agresti, (2002); indica que esta distribución se mantiene para cualquiera de los tres esquemas de muestreo utilizados (multinomial, Poisson y producto de multinomiales) para generar la tabla de contingencia  $\mathbf{y}$ .

En el apéndice 9.1. se muestra la deducción de esta función de probabilidad para el caso de muestreo producto de multinomiales.



#### **5.4.1. Medida de discrepancia**

Sea  $D: \Gamma \rightarrow \mathfrak{R}$  una función de variable real, tal que  $D(\mathbf{y})$  le asigna un número real a cada tabla  $\mathbf{y}$  que pertenezca al conjunto de referencia  $\Gamma$ , de tal forma que la tabla  $\mathbf{y}_1$  es al menos tan extrema que la tabla  $\mathbf{y}_2$ ; siempre que  $D(\mathbf{y}_1) \geq D(\mathbf{y}_2)$ , (Cytel Software, 2001 y 2003; Agresti y Wackerly; 1977); es decir, la medida de discrepancia permite ordenar las tablas de contingencia  $\mathbf{y}$  del conjunto de referencia  $\Gamma$  y en consecuencia identificar a aquellas más extremas, de tal manera que puedan probarse hipótesis estadísticas basadas en tales medidas de discrepancia.

#### **5.4.2. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$ simplemente ordenadas**

Las pruebas que se mostrarán en esta sección son aplicables a tablas de contingencia de dimensión  $R \times C$  en las cuales las categorías en las filas están medidas en escala nominal y las categorías de las columnas en escala al menos ordinal. En su aplicación más común, buscan comparar  $r$  tratamientos o poblaciones donde cada uno de ellos genera una respuesta categórica al menos ordinal.

Los datos están clasificados en  $c$  categorías, en las cuales se agrupan las frecuencias para cada una de las  $r$  poblaciones y se disponen en una tabla de

contingencia de dimensión  $R \times C$  como se mostró en la tabla 5. Las  $c$  categorías de las columnas están ordenadas pero las  $r$  poblaciones de las filas no lo están.

Supóngase que hay  $m_{i+}$  individuos, sujetos o elementos en la población  $i$  donde cada uno de estos genera una respuesta multinomial de tal manera que cae en alguna de las  $c$  categorías ordenadas. Sean las  $m_{i+}$  respuestas de la población  $i$  y sea  $\Pi_i = (\pi_{i1}, \pi_{i2}, \dots, \pi_{ic})$  el vector constituido por los parámetros multinomiales de la población  $i$ ;  $i = 1, 2, \dots, r$ ; entonces la hipótesis nula queda especificada como:

$$H_0 : \Pi_1 = \Pi_2 = \dots = \Pi_r$$

La hipótesis alternativa es que al menos un par de los vectores de parámetros multinomiales presenten componentes o proporciones poblacionales diferentes. Es necesario hacer notar que aunque  $H_0$  sea rechazada, las pruebas que se estudiarán en esta sección no especifican cuales poblaciones son estadísticamente diferentes.

Lo que se desea es probar si los  $r$  tratamientos o poblaciones presentan igual respuesta para las  $c$  categorías ordenadas o si hay al menos un par de poblaciones o tratamientos con respuestas diferentes, (Cytel Software, 2001 y 2003). Estas pruebas son denominadas frecuentemente en la literatura pertinente a la estadística no paramétrica como pruebas para  $r$  poblaciones independientes

(Siegel y Castellan, 1988); o como pruebas de rangos para varias muestras independientes, (Conover, 1971).

En esta sección se estudiarán las pruebas de Kruskal-Wallis, marcas normales (*normal scores*), marcas de Savage (*Savage scores*) y el análisis de varianza de una vía con marcas arbitrarias (*one-way ANOVA with arbitrary scores*). Para estas pruebas las medidas de discrepancia o estadísticos de prueba presentan la siguiente estructura funcional:

$$D(\mathbf{y}) = (S_w)^{-2} \sum_{i=1}^r [T_i - E(T_i)]^2 / m_{i+} \quad (2)$$

donde

$$T_i = \sum_{j=1}^c w_j y_{ij}$$

$$S_w^2 = \frac{1}{N-1} \sum_{j=1}^c n_{+j} (w_j - \bar{w})^2$$

$$\bar{w} = \frac{1}{N} \sum_{j=1}^c w_j n_{+j}$$

y además

- $w_j$  son las marcas (*scores*) de las columnas,
- $T_i$  es la suma observada de las marcas para todos los elementos o sujetos pertenecientes a la población  $i$ ,

- $E(T_i)$  es la esperanza de los  $T_i$  bajo la hipótesis nula,
- $S_w^2$  es la varianza muestral de las marcas.

#### 5.4.2.1. Prueba de Kruskal-Wallis

Es la prueba no paramétrica más popular para comparar  $r$  poblaciones no ordenadas; es una generalización de la prueba de suma de rangos de Wilcoxon-Mann-Whitney para dos muestras independientes. Para la tabla de contingencia observada,  $\mathbf{x}$ , el estadístico de prueba de Kruskal-Wallis, denotado por  $KW(\mathbf{x})$  está dado por

$$KW(\mathbf{x}) = \frac{12}{N(N+1) \left[ 1 - \lambda / (N^3 - N) \right]} \sum_{i=1}^r [t_i - m_{i+} (N+1)/2]^2 / m_{i+}$$

donde

$$t_i = \sum_{j=1}^c w_j x_{ij}$$

lo cual constituye la suma de los rangos medios en la población  $i$ , y las marcas  $w_j$  para esta prueba están dadas por

$$w_j = n_{+1} + n_{+2} + \dots + n_{+j-1} + \frac{1}{2}(n_{+j} + 1); \text{ para } j = 1, 2, \dots, c$$

y  $\lambda$  representa al factor de corrección por la presencia de empates, el cual se define por

$$\lambda = \sum_{i=1}^c (n_{+j}^3 - n_{+j})$$

Esta prueba es muy conocida y utilizada, sobre todo cuando no se puede asumir distribución alguna para las variables bajo estudio, excepto que provengan de poblaciones con distribuciones continuas. (Cytel Software, 2001 y 2003; Hollander y Wolfe, 1999).

Si bien para datos continuos esta prueba llega mostrar hasta una eficiencia del 98% en relación al análisis de varianzas de una vía de clasificación con datos homocedásticos, en datos categorizados tal eficiencia no aplica, lo que trae como consecuencia que existan otras pruebas más adecuadas para tales situaciones, (Cytel Software, 2003; Permutt y Berger, 2000).

#### 5.4.2.2. Prueba de marcas normales

Es una alternativa a la prueba de Kruskal-Wallis para comparar  $r$  poblaciones no ordenadas y está basada en los estadísticos de orden de la distribución normal.

El estadístico observado para esta prueba denotado por  $NO(\mathbf{x})$ , se calcula utilizando la expresión (2), página 41; donde las marcas vienen dadas por los estadísticos de orden esperados de la distribución normal estándar.

Así, para esta prueba, las marcas de las columnas  $w_j$  se definen como

$$w_j = \frac{1}{n_{+j}} \left[ \sum_{l=v_{j-1}+1}^{v_j} \Phi^{-1} \left( \frac{l}{N+1} \right) \right]; \text{ para } j = 1, 2, \dots, c$$

donde  $v_j = n_{+1} + n_{+2} + \dots + n_{+j}$ ;  $j = 1, 2, \dots, c$ ; y  $\Phi^{-1}(\alpha)$  es el percentil  $100\alpha\%$  de la distribución normal estándar, (Cytel Software, 2001 y 2003).

#### 5.4.2.3. Prueba de marcas de Savage

Es también una prueba alternativa a la prueba de Kruskal-Wallis para comparar  $r$  poblaciones no ordenadas y está basada en los estadísticos de orden de la distribución exponencial; por ello es la prueba más adecuada cuando se tienen razones para suponer que los datos provienen de poblaciones exponenciales. El estadístico de prueba para la tabla de contingencia observada  $\mathbf{x}$  denotado por  $SA(\mathbf{x})$  se calcula utilizando la expresión (2), página 41; pero considerando las marcas para las columnas definidas por

$$w_j = \frac{1}{n_{+j}} \left[ \sum_{l=v_{j-1}+1}^{v_j} \sum_{k=1}^l \left( \frac{1}{N-k+1} \right) \right] - 1; \text{ para } j = 1, 2, \dots, c$$

donde  $v_j = n_{+1} + n_{+2} + \dots + n_{+j}$ ;  $j = 1, 2, \dots, c$ ; (Cytel Software, 2001 y 2003).

#### 5.4.2.4. Análisis de varianza de una vía con marcas arbitrarias

Las tres pruebas vistas hasta ahora en esta sección solo se diferencian en la elección de las marcas de las columnas  $w_j$ ; así, cada una de estas pruebas es una particularización de una familia de pruebas cuyas marcas provienen de un

conjunto  $W$  de marcas posibles, esto es una consecuencia de lo general que es la definición del estadístico de prueba (2), tal generalidad deja abierta la posibilidad al investigador de proponer sus propias marcas, las cuales no necesariamente coincidan con las correspondientes a las pruebas vistas anteriormente. Las pruebas desarrolladas a partir de este razonamiento se denominan genéricamente análisis de varianza de una vía con marcas arbitrarias, puesto que las marcas son elección del investigador mismo, (Cytel Software, 2001 y 2003).

Graubard y Korn, (1987); estudiaron el comportamiento de pruebas para tablas de contingencia  $2 \times C$  utilizando diferentes tipos de marcas; específicamente rangos medios, puntos medios de las categorías y valores arbitrarios igualmente espaciados; ellos encontraron que el comportamiento de las tres pruebas generadas por estas marcas fue diferente, reflejándose éste en los  $p$ -valores exactos correspondientes a cada una de ellas. Concluyeron que cuando la tabla de contingencia es muy desbalanceada; es decir, la distribución de los totales de columnas no es muy uniforme, no necesariamente la elección de las marcas de rangos medios genera una prueba adecuada, ellos se inclinaron más bien hacia la elección del punto medio de la categoría como marca más adecuada; sin embargo, advierten que es bueno elegir las marcas antes de obtener el  $p$ -valor; es decir, no elegir las marcas que den resultados convenientes al investigador después de obtenido el  $p$ -valor, sino escoger aquellas marcas que reflejen razonablemente el

comportamiento de los datos. Adicionalmente, presentan una tabla donde muestran las marcas más utilizadas en las pruebas más comunes.

Más tarde, Permutt y Berger, (2000); revisaron el trabajo de Graubard y Korn y concluyeron que para el ejemplo desarrollado por éstos, estaban de acuerdo con la falta de precisión en el análisis llevado a cabo utilizando a los rangos medios de Wilcoxon-Mann-Whitney como marcas de columnas, pero que sin embargo, pudieron haber incluido en el estudio a la prueba de marcas de Savage, debido a la asimetría de la respuesta presente a lo largo de las columnas; comentan además, que dada esta misma asimetría no recomiendan el uso de la prueba de marcas normales en este estudio. Este punto de vista ratifica lo sugerido por Graubard y Korn, (1987); en lo concerniente a que las marcas de las columnas deben elegirse antes de obtener el  $p$ -valor correspondiente a la prueba utilizada, y que esta elección debe estar de acuerdo con la naturaleza de los datos.

Para las cuatro pruebas propuestas, el  $p$ -valor exacto está dado por

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

Asintóticamente todas las pruebas vistas en esta sección convergen a la distribución  $\chi^2$  con  $(r-1)$  grados de libertad y el  $p$ -valor asintótico está dado por

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



sin embargo, la tasa de convergencia de alguna prueba depende de la elección de las marcas de las columnas, en general, la tasa de convergencia será menor si las marcas están desigualmente espaciadas, (Cytel Software, 2001 y 2003).

Por otra parte, las tablas de contingencia de dimensión  $2 \times C$  constituyen un caso particular muy importante, y se pueden dividir en:

- El estudio de dos poblaciones multinomiales ordenadas.
- El estudio de  $C$  poblaciones binomiales.

Si se estudian dos poblaciones multinomiales ordenadas, las pruebas contempladas en esta sección les son aplicables sin modificación alguna, por cuanto que las categorías de columnas en ambos casos están ordenadas y no así las dos categorías de filas correspondientes a las poblaciones bajo estudio; la separación entre ambos métodos presente en la literatura obedece simplemente al hecho de que la prueba de suma de rangos de Wilcoxon-Mann-Whitney para tablas de contingencia  $2 \times C$  es un caso particular de la prueba de Kruskal-Wallis para tablas de contingencia de dimensión  $R \times C$  y que el análisis de varianza con marcas arbitrarias para las tablas de contingencia de dimensión  $R \times C$  en este caso se denomina simplemente prueba de permutación con marcas arbitrarias; las otras dos pruebas contempladas, las de marcas normales y marcas de Savage, reciben el mismo nombre para este tipo de tablas, (Cytel Software, 2001 y 2003).

Para evitar confusiones, en el presente trabajo se llamará prueba de Kruskal-Wallis indistintamente a la prueba de Kruskal-Wallis propiamente dicha y a la prueba de Wilcoxon-Mann-Whitney y ANAVAR con marcas arbitrarias a cualquier prueba de permutación con marcas asignadas arbitrariamente por el investigador, sin importar la dimensión de la tabla de contingencia considerada.

En el caso de  $C$  poblaciones binomiales, las pruebas desarrolladas para este caso difieren de las estudiadas en esta sección tanto en el fundamento teórico como en los estadísticos de prueba utilizados, razón por la cual no se incluyen en este estudio, para mayor información acerca de estas pruebas, ver Good, (2000); y Cytel Software, (2001 y 2003).

#### ***5.4.3. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$ doblemente ordenadas***

Las pruebas que se estudiarán en esta sección son aplicables a tablas de contingencia de dimensión  $R \times C$  en las cuales tanto las categorías de las filas como las categorías de las columnas están ordenadas, es decir, tanto las filas como las columnas de la tabla de contingencia están medidas en escala al menos ordinal.

En las secciones anteriores se revisaron aspectos teóricos relacionados con pruebas que están diseñadas para tablas de contingencia en las que sólo la variable respuesta está ordenada; tal estructura impide el diseño de pruebas de hipótesis más específicas, puesto que las pruebas sólo verifican si existe igual respuesta en  $r$

poblaciones, pero no se puede estudiar si existe algún tipo de tendencia además de la homogeneidad de los vectores de probabilidad multinomial de las  $r$  filas. El doble ordenamiento de ambas variables para las tablas de contingencia de esta sección permite formular hipótesis en las cuales se estudia si existe algún tipo de tendencia entre la variable explicativa  $X$  y la variable respuesta  $Y$ . Para ilustrar esto, supóngase que cada respuesta cae dentro de una de  $c$  categorías ordinales de acuerdo a una distribución multinomial. Sea  $m_{i+}$  las respuestas de la población  $i$  que caen en las  $c$  categorías ordenadas y sea  $\Pi_i = (\pi_{i1}, \pi_{i2}, \dots, \pi_{ic})$  el vector de probabilidades multinomiales;  $i = 1, 2, \dots, r$ . Entonces, la hipótesis nula es

$$H_0 : \Pi_1 = \Pi_2 = \dots = \Pi_r$$

Para especificar la hipótesis alternativa, sea

$$\gamma_{ij} = \sum_{l=1}^j \pi_{il}$$

Como las filas están ordenadas, es posible definir una hipótesis alternativa unilateral de la forma

$$H_1 : \gamma_{1j} \geq \gamma_{2j} \geq \dots \geq \gamma_{rj} \quad \text{ó} \quad H'_1 : \gamma_{1j} \leq \gamma_{2j} \leq \dots \leq \gamma_{rj}$$

para  $i = 1, 2, \dots, r$ ; con al menos una desigualdad estricta en algún nivel  $j$ .

Las pruebas que se revisarán en esta sección son la prueba de Jonckheere-Terpstra y de asociación lineal-lineal; ambas permiten probar hipótesis unilaterales

del tipo  $H_1$  o  $H_1'$ , donde para valores positivos grandes se falla a favor de  $H_1$  y para valores negativos grandes se falla a favor de  $H_1'$ .

Así mismo, se puede proponer una hipótesis alternativa bilateral la cual será válida si  $H_1$  o  $H_1'$  se mantienen.

La hipótesis alternativa  $H_1$  implica que en la medida que la variable  $X$  se incrementa de  $i$  a  $(i+1)$  la probabilidad de que la respuesta caiga en la categoría  $(j+1)$  en vez de  $j$  aumenta; la hipótesis alternativa  $H_1'$  establece lo contrario.

#### 5.4.3.1. Prueba de Jonckheere-Terpstra

La prueba de Jonckheere-Terpstra ordena las tablas en  $\Gamma$  de acuerdo al estadístico de Jonckheere-Terpstra el cual es en realidad la suma de  $r(r-1)/2$  estadísticos de Wilcoxon-Mann-Whitney. Específicamente para la tabla de contingencia observada  $\mathbf{x}$  de dimensión  $R \times C$ , el estadístico de prueba no normalizado, denotado por  $JT(\mathbf{x})$ , está dado por

$$JT(\mathbf{x}) = \sum_{i=2}^r \sum_{j=1}^{i-1} \sum_{k=1}^c \left[ w_{ijk} y_{ik} - m_{i+} (m_{i+} + 1) / 2 \right]$$

donde los valores  $w_{ijk}$  son las marcas de Wilcoxon correspondientes a una tabla de dimensión  $2 \times C$  formada por las filas  $i$  y  $j$  de la tabla  $R \times C$ . Así, para  $k = 1, 2, \dots, c$

$$w_{ijk} = \left[ (x_{i1} + x_{j1}) + \dots + (x_{i,k-1} + x_{j,k-1}) + (x_{ik} + x_{jk} + 1) \right] / 2$$

Bajo la hipótesis nula de que las  $r$  poblaciones poseen los mismos vectores  $\Pi_i$  de probabilidades multinomiales, el estadístico de Jonckheere-Terpstra tiene como media

$$E[JT(\mathbf{X})] = \frac{1}{4} \left( N^2 - \sum_{i=1}^r m_{i+}^2 \right)$$

y varianza

$$\begin{aligned} Var[JT(\mathbf{X})] = & \frac{1}{72} \left[ N(N-1)(2N+5) - \sum_{i=1}^r m_{i+} (m_{i+} - 1)(2m_{i+} + 5) - \sum_{j=1}^c n_{+j} (n_{+j} - 1)(2n_{+j} + 5) \right] \\ & + \frac{1}{36N(N-1)(N-2)} \left[ \sum_{i=1}^r m_{i+} (m_{i+} - 1)(m_{i+} - 2) \right] \left[ \sum_{j=1}^c n_{+j} (n_{+j} - 1)(n_{+j} - 2) \right] \\ & + \frac{1}{8N(N-1)} \left[ \sum_{i=1}^r m_{i+} (m_{i+} - 1) \right] \left[ \sum_{j=1}^c n_{+j} (n_{+j} - 1) \right] \end{aligned}$$

luego, la forma estandarizada del estadístico de Jonckheere-Terpstra es

$$Z_{JT(\mathbf{X})} = \frac{JT(\mathbf{X}) - E[JT(\mathbf{X})]}{\sqrt{Var[JT(\mathbf{X})]}}$$

el cual para muestras grandes converge a la distribución normal estándar, (Hollander y Wolfe, 1999; Cytel Software, 2001 y 2003)

#### 5.4.3.2. Prueba de asociación lineal-lineal

La prueba de asociación lineal-lineal ordena las tablas en  $\Gamma$  de acuerdo a los estadísticos de rangos lineales. Así, si la tabla observada es  $\mathbf{x}$ , el estadístico no estandarizado, denotado por  $LL(\mathbf{x})$  viene dado por

$$LL(\mathbf{x}) = \sum_{i=1}^r \sum_{j=1}^c u_i v_j y_{ij}$$

donde  $u_i; i = 1, 2, \dots, r$ ; son las marcas arbitrarias para las filas y  $v_j; j = 1, 2, \dots, c$ ; son las marcas arbitrarias para las columnas. Bajo la hipótesis nula de no interacción entre filas y columnas, el estadístico de asociación lineal-lineal tiene como media

$$E[LL(\mathbf{X})] = N^{-1} \left( \sum_{i=1}^r u_i m_{i+} \right) \left( \sum_{j=1}^c v_j n_{+j} \right)$$

y varianza

$$Var[LL(\mathbf{X})] = (N-1)^{-1} \left[ \sum_{i=1}^r u_i^2 m_{i+} + \frac{1}{N} \left( \sum_{i=1}^r u_i m_{i+} \right)^2 \right] \left[ \sum_{j=1}^c v_j^2 n_{+j} + \frac{1}{N} \left( \sum_{j=1}^c v_j n_{+j} \right)^2 \right]$$

luego, la forma estandarizada del estadístico de asociación lineal-lineal es

$$Z_{LL(\mathbf{X})} = \frac{LL(\mathbf{X}) - E[LL(\mathbf{X})]}{\sqrt{Var[LL(\mathbf{X})]}}$$

el cual para muestras grandes converge a la distribución normal estándar, (Cytel Software, 2001 y 2003; Agresti, 2002)

Para ambas pruebas, sea  $Z_{D(y)}$  la forma estandarizada del estadístico de prueba o medida de discrepancia para una tabla de contingencia  $y \in \Gamma$  y sea  $Z_{D(x)}$  el valor de la forma estandarizada del estadístico de prueba para la tabla de contingencia observada  $x$ . Entonces el  $p$ -valor exacto unilateral para ambas pruebas está dado por

$$p_1 = \begin{cases} \Pr[Z_{D(y)} \geq Z_{D(x)}] & \text{si } Z_{D(x)} > 0 \\ \Pr[Z_{D(y)} \leq Z_{D(x)}] & \text{si } Z_{D(x)} \leq 0 \end{cases}$$

y el  $p$ -valor exacto bilateral para ambas pruebas está dado por

$$p_2 = \Pr\left[\left|Z_{D(y)}\right| \geq \left|Z_{D(x)}\right|\right]$$

Por otra parte, el  $p$ -valor asintótico unilateral para estas dos pruebas está dado por

$$\tilde{p}_1 = \min\left\{\Phi\left[Z_{D(x)}\right], 1 - \Phi\left[Z_{D(x)}\right]\right\}$$

y el  $p$ -valor asintótico bilateral para ambas pruebas está dado por

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

Donde  $\Phi[Z_{D(x)}]$  es el valor acumulado de la distribución normal estándar para el valor estandarizado observado  $Z_{D(x)}$ .

Si bien ambos estadísticos convergen a la distribución normal estándar para el caso de muestras grandes; en el caso de muestras pequeñas no se han llevado a cabo estudios acerca del comportamiento de la función de potencia para estas dos pruebas, (Cytel Software, 2001 y 2003); así pues, ambas pueden ser utilizadas indistintamente para este tipo de tablas, puesto que no hay evidencias para recomendar una en lugar de otra. Sin embargo, la prueba de asociación lineal-lineal pudiera tener cierta ventaja sobre la prueba de Jonckheere-Terpstra, puesto que aquella permite el uso de marcas arbitrarias elegidas por el investigador y que no necesariamente coincidan con los rangos de Wilcoxon-Mann-Whitney, así pues, las consideraciones hechas en la sección anterior acerca de la elección de los rangos para el análisis de varianza de una vía con marcas arbitrarias son válidas también para la prueba de asociación lineal-lineal, (Cytel Software, 2001 y 2003; Agresti, 2002).

Agresti, Mehta y Patel, (1990); comentan que aún no se han desarrollado guías para la aplicación de las aproximaciones asintóticas, es decir, no está claro a partir de que punto se pueden considerar a la pruebas de permutación y a su aproximaciones asintóticas como equivalentes.



Por otra parte, si las marcas de filas y columnas son los datos mismos, la prueba de asociación lineal-lineal coincide con la prueba de significación del coeficiente de correlación lineal de Pearson y si en cambio, las marcas son reemplazadas por los rangos medios de Wilcoxon, la prueba coincide con la prueba de significación del coeficiente de correlación de Spearman, (Mehta y Patel, 1997).

### **5.5. Aspectos computacionales sobre pruebas de permutación en tablas de contingencia**

La presente sección trata los aspectos computacionales para llevar a cabo pruebas exactas en tablas de contingencia. En la actualidad los algoritmos desarrollados para tal fin realizan los cálculos correspondientes a partir de dos enfoques diferentes:

- Enumeración completa del conjunto de referencia  $\Gamma$  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 halla todas las tablas que presenten los mismos totales marginales de filas y columnas que la tabla observada; es decir, aquellas que constituyen el conjunto de referencia  $\Gamma$ . A partir de este conjunto se calcula el  $p$ -valor exacto como la suma de todas las probabilidades

hipergeométricas de las tablas presentes en  $\Gamma$  cuya medida de discrepancia sea tan o más extrema que la de la tabla observada.

En la década de los 70 aparecieron los primeros algoritmos para realizar pruebas exactas en tablas de contingencia, los cuales fueron escritos en lenguaje FORTRAN y básicamente realizaban la enumeración completa del conjunto de referencia  $\Gamma$  para construir la distribución del estadístico de prueba con el fin de calcular el  $p$ -valor exacto.

Los algoritmos desarrollados durante este período presentaban como desventaja el intenso requerimiento computacional necesario para enumerar todo el conjunto de referencia  $\Gamma$  y en consecuencia su uso estaba limitado a tablas de contingencia de dimensión y tamaños de muestra pequeños, (Thomas, 1971; Baptista y Pike, 1977). Los investigadores recomendaban frecuentemente calcular previamente el número total de tablas de contingencia presentes en el conjunto de referencia  $\Gamma$  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  $\Gamma$ .

Dado que el tamaño del conjunto de referencia  $\Gamma$  crece geoméricamente con la dimensión de la tabla y el tamaño de muestra, (Agresti, 2002; Cytel Software,

2001 y 2003; Michaelides, 1997); el uso de tales algoritmos estuvo sumamente limitado.

Mehta y Patel, (1980); propusieron esquematizar al conjunto de referencia  $\Gamma$  como una estructura de red, donde todos los posibles caminos dentro de este retículo 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; posteriormente, Mehta y Patel, (1983); utilizaron tal estructura y el algoritmo correspondiente para obtener los  $p$ -valores exactos de la prueba de Freeman-Halton sin tener que calcular el valor de la medida de discrepancia para todas las tablas posibles, sino solamente para aquellas tan o más extremas que la correspondiente a la tabla observada, las cuales se utilizan para calcular el  $p$ -valor exacto; por cuanto que en la estructura de red tales tablas se pueden identificar fácilmente; evidentemente, 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.

Posteriormente tal algoritmo ha sido utilizado para calcular  $p$ -valores exactos en diferentes pruebas y esquemas de tablas de contingencia; entre los que se pueden citar:

- Mehta, Patel y Gray, (1985); utilizaron el algoritmo de red para 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 calcular *p*-valores exactos en tablas de contingencia cuadradas cuasi-simétricas con el fin de 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); cuyo algoritmo enumera las

tablas del conjunto de referencia  $\Gamma$  mediante métodos iterativos de la manera siguiente:

- (i) Enumerar todas las tablas de dimensión  $(R-1) \times C$  con marginales fijas,
- (ii) Enumerar las tablas de dimensión  $2 \times C$  con marginales fijas y por último,
- (iii) Enumerar las tablas  $2 \times 2$  con marginales fijas.

Baglivo, Olivier y Pagano, (1988); desarrollaron un algoritmo basado en un principio diferente a los citados anteriormente; el cual consiste en 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 ambas subtablas, de manera que este algoritmo híbrido requiere menos intensidad de procesamiento computacional, lo cual lo hace más rápido que los otras metodologías, 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 para tablas de contingencia conocido como Monte Carlo clásico. Este ensayo selecciona tablas aleatoriamente del conjunto de referencia  $\Gamma$  con probabilidad hipergeométrica múltiple. Bajo este esquema de simulación es suficiente una muestra aleatoria de 17000 tablas de contingencia con el objeto de construir 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 el cociente 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  $2 \times 2$ ; sin embargo, ellos seleccionaron las tablas utilizando distribuciones de probabilidad diferentes a la hipergeométrica múltiple, tales como log-normal, uniforme, exponencial y distribuciones discretas dicotómicas. 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 identifica previamente en la red a aquellos caminos que corresponden a las tablas de contingencia tan o más extremas que la tabla observada, constituyendo así el conjunto de referencia restringido  $\Gamma^*$ ; posteriormente selecciona tablas de este conjunto  $\Gamma^*$  según su importancia para reducir la varianza del estimador del  $p$ -valor exacto. Es decir, este esquema de muestreo sólo selecciona tablas en  $\Gamma^*$  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 para 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 propusieron 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  $\Gamma$  que pertenezcan al conjunto de referencia reducido  $\Gamma'$ , constituido por las tablas de contingencia que además de tener marginales de filas y columnas fijas presenten las mismas frecuencias conjuntas  $y_{ii}; i = 1, 2, \dots, r; r = c$ ; en la diagonal principal de la tabla de contingencia observada; este conjunto de referencia reducido  $\Gamma'$  es mucho más pequeño que  $\Gamma$ , tal como lo muestran McDonald, De Roure y Michaelides, (1999). Es importante señalar que este algoritmo, luego de identificar al conjunto de referencia reducido  $\Gamma'$ , selecciona las tablas de este conjunto con probabilidad hipergeométrica múltiple de forma similar al Monte Carlo clásico.

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; básicamente, 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). Smith, Forster y McDonald, (1996); aplicaron estas metodologías para estimar 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  $R \times C$  simple y doblemente ordenadas con su correspondiente equivalente asintótico, con el objeto 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 $R \times C$ simplemente ordenadas*

*Ejemplo 1:* La tabla de contingencia 7 fue construida a partir de los resultados de un ensayo clínico llevado a cabo por Khatri y colaboradores, (1985); citado por Bajorski y Petkau, (1999); en el cual, un grupo de pacientes tratados por esclerosis múltiple crónica y progresiva recibieron un tratamiento consistente en una plasmaféresis y un placebo consistente en una falsa plasmaféresis y fueron tratados durante 20 semanas. La respuesta se resumió en una tabla de contingencia de dimensión  $2 \times 3$  simplemente ordenada, donde las filas representan al tratamiento y al placebo y las columnas presenta el estado de la respuesta de los pacientes según hallan mostrado mejoría o no al final del lapso de estudio.

**Tabla 7.** Estado de pacientes con esclerosis múltiple crónica y progresiva al cabo de 20 semanas de tratamiento.

ESTADO DEL PACIENTE AL FINAL DEL TRATAMIENTO				
	Peor	Igual	Mejor	Total
Tratamiento	1	11	14	26
Control	3	18	8	29
Total	4	29	22	55

Fuente: Bajorski y Petkau, 1999.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 8.

**Tabla 8.** Resultados de los  $p$ -valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 1.

PRUEBAS DE SIGNIFICACIÓN				
	Kruskal-Wallis	Normal	Savage	ANAVAR
$p$ -valor exacto	0,04461862	0,05757899	0,04740451	0,07402506
$p$ -valor asintótico	0,04329079	0,04591694	0,04403713	0,04685513
$\Delta$	0,00132783	0,01166205	0,00336738	0,02716993

Para la tabla de contingencia 7, con tamaño de muestra moderadamente pequeño  $N = 55$ ; y totales marginales de filas balanceados, se observa en la tabla 8 que de las cuatro pruebas de significación aplicadas, solamente la prueba de marcas normales y el ANAVAR con marcas arbitrarias presentaron valores  $\Delta$

superiores al umbral establecido  $\Delta^* = 0,01$ ; en cuyo caso se utilizaría para sus análisis a las metodologías exactas siguiendo la recomendación de Berger, (2000); mencionada anteriormente, la cual establece que si el valor de las diferencias entre los  $p$ -valores exacto y asintótico  $\Delta$  es inferior al umbral  $\Delta^*$  se pueden utilizar métodos asintóticos y en caso contrario métodos exactos. Por otra parte, en las pruebas de Kruskal-Wallis y marcas de Savage los valores  $\Delta$  resultaron menores al umbral  $\Delta^*$ , lo cual llevaría a basar la decisión en los métodos asintóticos.

Se observa además, que en aquellas pruebas en las cuales  $\Delta$  es superior al umbral  $\Delta^*$ ; es decir, en la prueba de marcas normales y el ANAVAR con marcas arbitrarias, las decisiones tomadas al nivel de significación  $\alpha$  de 5% hubieran sido diferentes al utilizar métodos exactos en vez de métodos asintóticos; es decir, no se hubiera podido rechazar la hipótesis nula de igualdad de los vectores de parámetros multinomiales de las filas  $\Pi_i$ ,  $i = 1, 2$ ; en cambio, se hubiera rechazado esta hipótesis nula si la decisión hubiera estado basada en los métodos asintóticos. Por otra parte, en aquellas dos pruebas en las que  $\Delta$  fue inferior al umbral  $\Delta^*$ ; es decir, las pruebas de Kruskal-Wallis y marcas de Savage, las decisiones tomadas al nivel de significación  $\alpha$  de 5% hubieran sido iguales tanto si se tomara el  $p$ -valor exacto como el  $p$ -valor asintótico; en este caso ambos hubieran indicado que hay diferencias entre los vectores de parámetros multinomiales del

tratamiento y el control; es decir, las dos poblaciones trinomiales no son homogéneas.

*Ejemplo 2:* La siguiente tabla de contingencia muestra el número de pacientes dados de alta vivos o muertos, de acuerdo a las unidades de volumen de émbolos presentados en aneurismas aórtico, (Bayly, Matthews, Dobson, Price y Thomas, 2001). Tal estudio provee una tabla de contingencia de dimensión 2x3, como se presenta en la tabla 9.

**Tabla 9.** *Pacientes dados de alta vivos o muertos, de acuerdo a las unidades de volumen de émbolos presentados en aneurismas aórtico.*

UNIDADES DE VOLUMEN DE ÉMBOLOS				
	Bajo	Medio	Alto	Total
De alta muertos	19	24	13	56
De alta vivos	261	319	175	755
Total	280	343	188	811

Fuente: Bayly, Matthews, Dobson, Price y Thomas, 2001.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 10.

**Tabla 10.** Resultados de los  $p$ -valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 2.

	PRUEBAS DE SIGNIFICACIÓN			
	Kruskal-Wallis	Normal	Savage	ANAVAR
$p$ -valor exacto	0,96252099	0,97063812	0,97679864	0,96356506
$p$ -valor asintótico	0,94494171	0,94724180	0,96766607	0,95945539
$\Delta$	0,01754928	0,02339632	0,00913257	0,00410967

Para la tabla de contingencia 9, cuyos totales marginales de filas son fuertemente desbalanceados y con tamaño de muestra grande  $N = 811$ ; se observa en la tabla 10 que para las pruebas de Kruskal-Wallis y de marcas normales las diferencias entre los  $p$ -valores exactos y asintóticos fueron superiores al umbral  $\Delta^*$ ; en cuyo caso, siguiendo la recomendación de Berger, (2000); deberían utilizarse métodos exactos para su análisis; mientras que en las pruebas de marcas de Savage y el ANAVAR con marcas arbitrarias, las diferencias entre los  $p$ -valores  $\Delta$  fueron inferiores al umbral  $\Delta^*$ ; en este caso se podrían utilizar métodos asintóticos.

También se observa que en las cuatro pruebas aplicadas e independientemente del método utilizado, los  $p$ -valores fueron muy cercanos a 1, lo que conduce a sospechar que la hipótesis nula de igualdad de los vectores de parámetros multinomiales en las poblaciones de las filas es en realidad cierta al nivel de significación  $\alpha$  de 5%.

Resulta interesante observar que a pesar de que el tamaño de muestra es tan grande,  $N = 811$ ; en dos de las pruebas, Kruskal-Wallis y marcas normales, no se presentó la convergencia deseada de los  $p$ -valores asintóticos hacia los  $p$ -valores exactos, ya que las diferencias  $\Delta$  observadas fueron superiores al umbral  $\Delta^*$ , pareciera ser entonces que el fuerte desbalanceo en los totales marginales de las filas pudiera estar afectando notablemente la tasa de convergencia de los  $p$ -valores para estas pruebas.

*Ejemplo 3:* McCullagh, (1980); citado por Nair, (1987); presenta los resultados de un estudio de la calidad de la visión en el ojo derecho en hombres y mujeres; tales resultados generaron una tabla de contingencia simplemente ordenada de dimensión  $2 \times 4$ , donde las categorías de las filas representan el sexo de los individuos considerados en el estudio y las columnas representan la calidad de la visión, ordenada de forma creciente, tal y como se presenta en la tabla 11.

**Tabla 11.** *Calidad de la visión del ojo derecho en hombres y mujeres.*

CALIDAD DE LA VISIÓN EN EL OJO DERECHO					
	La más alta (1)	2	3	La más baja (4)	Total
Hombres	32	56	48	64	200
Mujeres	22	66	60	52	200
Total	54	122	108	116	400

Fuente: Nair, 1987.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 12.

**Tabla 12.** Resultados de los  $p$ -valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 3.

	PRUEBAS DE SIGNIFICACIÓN			
	Kruskal-Wallis	Normal	Savage	ANAVAR
$p$ -valor exacto	0,83472789	0,92936572	0,41392255	0,96123026
$p$ -valor asintótico	0,83473516	0,92937596	0,41552019	0,92249355
$\Delta$	0,00000727	0,00001024	0,00159764	0,03873671

Para tabla de contingencia 11, la cual presenta un tamaño de muestra grande  $N = 400$ ; y los totales marginales de filas son balanceados, se observa en la tabla 12 que en las pruebas de Kruskal-Wallis, marcas normales y marcas de Savage la convergencia de los  $p$ -valores asintóticos a los  $p$ -valores exactos es casi exacta; es decir, los valores  $\Delta$  encontrados fueron sumamente pequeños y mucho menores que el umbral  $\Delta^*$ . En cambio, para la prueba del ANAVAR con marcas arbitrarias se encontró que el valor  $\Delta$  fue superior al umbral  $\Delta^*$ ; de hecho, este valor  $\Delta$  es casi cuatro veces mayor que el umbral  $\Delta^*$ ; pudiera pensarse ante esta situación que ocurre lo que comentan Ivanova y Berger, (2001); en cuanto al uso de marcas enteras; es decir, cuando las marcas utilizadas son números enteros, ellos mostraron que las pruebas devenían más conservadoras y por ende menos



potentes que aquellas pruebas en las cuales las marcas utilizadas no fueran enteras. Esta situación pudiera estar afectando la convergencia del  $p$ -valor asintótico al  $p$ -valor exacto para esta prueba.

*Ejemplo 4:* La siguiente tabla de contingencia reportada por Edwardes, (1997); muestra los resultados de un estudio donde se desea verificar si el estilo de vida, rural o urbano, está relacionado o no con el crecimiento de tumores pulmonares durante un período de dos meses, tales resultados se muestran en la tabla 13.

**Tabla 13.** *Extensión del crecimiento del tumor pulmonar según el estilo de vida del paciente.*

EXTENSIÓN DEL TUMOR						
	Indetectable	Pequeño	Mediano	Grande	Extremo	Total
Rural	12	5	2	3	2	24
Urbana	16	27	8	13	8	72
Total	28	32	10	16	10	96

Fuente: Edwardes, 1997.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 14.

**Tabla 14.** Resultados de los  $p$ -valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 4.

	PRUEBAS DE SIGNIFICACIÓN			
	Kruskal-Wallis	Normal	Savage	ANAVAR
$p$ -valor exacto	0,05707982	0,05751223	0,21367362	0,13648232
$p$ -valor asintótico	0,05596531	0,05813933	0,20626713	0,11472647
$\Delta$	0,00111451	0,00062710	0,00740649	0,02175585

Para la tabla de contingencia 13, cuyo tamaño de muestra puede considerarse moderado  $N = 96$  ; y con totales marginales de las filas desbalanceados, se observa en la tabla 14 que en tres de las cuatro pruebas de significación aplicadas, Kruskal-Wallis, marcas normales y marcas de Savage; las diferencias entre los  $p$ -valores exactos y asintóticos  $\Delta$  fueron inferiores al umbral  $\Delta^*$ , lo cual indica que hubo convergencia de los  $p$ -valores asintóticos a los  $p$ -valores exactos y por ende, pudieran utilizarse metodologías asintóticas. En cambio, en el ANAVAR con marcas arbitrarias, la diferencia  $\Delta$  fue más del doble del umbral  $\Delta^*$ ; pudiera argumentarse que un fenómeno similar al mostrado en el ANAVAR del ejemplo anterior también ocurre este caso, debido a que las marcas utilizadas en esta prueba son enteras y al desbalanceo existente en los totales marginales de las filas de la tabla de contingencia, lo cual parece disminuir la tasa de convergencia en esta prueba.

Por otra parte, si bien en las cuatro pruebas aplicadas la decisión a tomar al nivel de significación  $\alpha$  de 5% es la misma; es decir, no rechazar la hipótesis nula de igualdad de los vectores de parámetros multinomiales entre los estilos de vida rural y urbano, los  $p$ -valores, tanto asintóticos como exactos, variaron mucho de una prueba a otra; así, en las pruebas de Kruskal-Wallis y de marcas normales éstos estuvieron alrededor de 6% mientras que en la prueba de marcas de Savage estuvieron alrededor de 20% y en el ANAVAR con marcas arbitrarias oscilaron entre 11 y 13%. Este comportamiento tan dispar mostrado entre las diferentes pruebas ratifica las observaciones encontradas en la literatura acerca de lo cuidadoso que debe ser el investigador al escoger la prueba a aplicar en un caso particular, ya que ésta debe corresponder con los supuestos subyacentes a las poblaciones bajo estudio, (Cytel Software, 2001 y 2003); ya que dependiendo de la forma distribucional de la población bajo estudio, las pruebas mostrarán un comportamiento diferente.

A pesar de que el tamaño de muestra no es muy grande,  $N = 96$ ; se observó una clara convergencia entre los  $p$ -valores exacto y asintótico para las pruebas de Kruskal-Wallis, marcas normales y marcas de Savage. Este comportamiento puede deberse a la dimensión de la tabla de contingencia, ya que esta consta de cinco categorías para las columnas, lo cual aumenta considerablemente el tamaño del

conjunto de referencia  $\Gamma$  de tablas de contingencia condicionadas y pudiera acelerar la convergencia de los  $p$ -valores asintóticos hacia los  $p$ -valores exactos.

*Ejemplo 5:* Mills y Graubard, (1987); citados por Graubard y Korn, (1987); presentan los resultados de un estudio sobre el consumo de alcohol por parte de madres durante los primeros tres meses de embarazo y la presencia de malformaciones congénitas en los órganos genitales de los recién nacidos; este estudio generó una tabla de contingencia simplemente ordenada de dimensión  $2 \times 5$  la cual se presenta en la tabla 15.

**Tabla 15.** Malformaciones congénitas en los órganos genitales de los recién nacidos según el consumo de alcohol de las madres durante los tres primeros meses de embarazo.

		CONSUMO DE ALCOHOL (Promedio de tragos·día <sup>-1</sup> )					
		0	< 1	1-2	3-5	≥ 6	Total
Malformaciones ausentes		17066	14464	788	126	37	32481
Malformaciones presentes		48	38	5	1	1	93
<b>Total</b>		17114	14502	793	127	38	32574

Fuente: Graubard y Korn, 1987.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  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  $\Delta$  correspondiente para los datos del ejemplo 5.

	PRUEBAS DE SIGNIFICACIÓN			
	Kruskal-Wallis	Normal	Savage	ANAVAR
$p$ -valor exacto	0,57492965	0,29148829	0,05137758	0,17896779
$p$ -valor asintótico	0,55330615	0,28109586	0,05152168	0,17639240
$\Delta$	0,02162350	0,01039243	0,00014410	0,00257539

Para la tabla de contingencia 15, la cual presenta un tamaño de muestra muy grande  $N = 32574$ ; y un fuerte desbalanceo en los totales marginales de las filas, siendo  $m_{1+} = 32481$  y  $m_{2+} = 93$ , para la ausencia y presencia de malformaciones congénitas en los órganos genitales de los recién nacidos, respectivamente; se observa en la tabla 16, que al contrario de lo que se esperaría para un tamaño de muestra tan grande, las diferencias  $\Delta$  en las pruebas de Kruskal-Wallis y de marcas normales fueron superiores al umbral  $\Delta^*$ , este comportamiento corrobora lo encontrado en la literatura acerca de que el desbalanceo afecta la tasa de convergencia de las pruebas asintóticas hacia las pruebas exactas, (Cytel Software, 2001 y 2003). En cambio, en la prueba de marcas de Savage la diferencia de los  $p$ -valores exacto y asintótico  $\Delta$  fue inferior al umbral  $\Delta^*$ ; esta clara convergencia puede deberse al hecho de que la distribución de los totales marginales de las columnas es decreciente y pudiera sospecharse que la distribución subyacente a la

población bajo estudio es similar a la distribución exponencial negativa; de hecho, Permutt y Berger, (2000); señalan que para esta tabla de contingencia; Graubard y Korn (1987); pudieron haber aplicado la prueba de marcas de Savage en su trabajo acerca de la elección de marcas en tablas de contingencia ordenadas, tal como se hizo en el presente informe, ya que la distribución marcadamente asimétrica y decreciente pudiera ajustarse a una distribución exponencial negativa. Tal similitud con esta distribución pudiera ser la causa de que la prueba que mostró mejor convergencia fue precisamente la prueba de marcas de savage, la cual tal como se afirma en Cytel Software, (2001 y 2003); es la más adecuada si las poblaciones bajo estudio presentan distribución exponencial negativa.

El ANAVAR con marcas arbitrarias, por otra parte, mostró una convergencia ligeramente menor que la prueba de marcas de Savage; de hecho, Graubard y Korn, (1987); para este ejemplo en particular recomendaron el uso de marcas enteras igualmente espaciadas en vez de la prueba de Kruskal-Wallis, debido a que las marcas de Wilcoxon-Mann-Whitney asignan valores muy bajos a las dos primeras categorías de la tabla de contingencia, que es precisamente donde se encuentran la mayoría de las observaciones; lo cual no refleja en esencia el comportamiento de la población subyacente; al contrario del ANAVAR con marcas arbitrarias que asigna igual peso a todas las categorías de las columnas. Graubard y Korn, (1987); concluyeron que las marcas de Wilcoxon-Mann-Whitney utilizadas

por la prueba de Kruskal-Wallis no son adecuadas para tablas de contingencia que presenten un fuerte desbalanceo en los totales marginales de las columnas y se inclinan hacia el uso de marcas igualmente espaciadas y de rangos medios de las columnas, en caso de que éstas sean cuantitativas.

Adicionalmente, se observa que los  $p$ -valores variaron mucho de una prueba a otra, siendo muy grandes en la prueba de Kruskal-Wallis y mostrando los menores valores en la prueba de marcas de Savage; así, el fuerte desbalanceo parece afectar el comportamiento de las diferentes pruebas y se ratifica la observación de que la prueba a utilizarse deber ser escogida cuidadosamente por el investigador basándose en la naturaleza de las poblaciones subyacentes, (Graubard y Korn, 1987; Cytel Software, 2001 y 2003).

Por último, cabe destacar que este ejemplo es muy citado en la literatura para ilustrar como la escogencia de las marcas y el desbalanceo de la tabla de contingencia afectan al comportamiento de pruebas exactas y asintóticas en el análisis de datos categorizados; este ejemplo ha sido citado por Graubard y Korn, (1987); Good, (2000); Permutt y Berger, (2000); Cytel Software, (2001 y 2003); y Agresti, (2002).

*Ejemplo 6:* La siguiente tabla de contingencia de dimensión  $2 \times 6$  presentada por Vecchia y colaboradores (1986); citada por Mehta, Patel y Senchaudhuri, (1988); muestra los resultados de un estudio caso control donde se relaciona el tiempo de

duración del hábito de fumar y la presencia de cáncer cervical, las categorías de las columnas representan el tiempo, en años, del hábito de fumar de las pacientes, los resultados se presentan en la tabla 17.

**Tabla 17.** Tiempo de duración del hábito de fumar y presencia de cáncer cervical.

	DURACIÓN DEL HÁBITO DE FUMAR (Años)						Total
	Nunca	< 10	10-19	20-29	30-39	≥ 40	
Cáncer cervical	155	10	19	25	11	10	230
Control	169	10	21	20	7	3	230
<b>Total</b>	324	20	40	45	18	13	460

Fuente: Mehta, Patel y Senchaudhuri, 1988.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  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  $\Delta$  correspondiente para los datos del ejemplo 6.

	PRUEBAS DE SIGNIFICACIÓN			
	Kruskal-Wallis	Normal	Savage	ANAVAR
$p$ -valor exacto	0,08446627	0,04932786	0,02482725	0,04170523
$p$ -valor asintótico	0,08441490	0,04943424	0,02520458	0,03867031
$\Delta$	0,00005137	0,00010638	0,00037733	0,00303492

Para la tabla de contingencia 17, la cual presenta un tamaño de muestra grande  $N = 460$  ; y totales marginales de filas balanceados y seis categorías



ordenadas en las columnas; se observa en la tabla 18 que para todas las pruebas aplicadas, las diferencias entre los  $p$ -valores exactos y asintóticos  $\Delta$  fueron bastante menores que el umbral  $\Delta^*$ ; este comportamiento apoya el comentario encontrado en Cytel Software, (2001 y 2003); donde se afirma que el tamaño del conjunto de referencia  $\Gamma$  de las tablas de contingencia condicionadas aumenta geométricamente con el tamaño de muestra y la dimensión de la tabla de contingencia observada, y que tal conjunto de referencia  $\Gamma$  es más grande en la medida que la muestra sea más balanceada; esta característica redundante en una más rápida convergencia entre los  $p$ -valores asintóticos y exactos, tal y como se aprecia en este ejemplo.

Resulta interesante señalar que si bien los  $p$ -valores asintóticos convergen adecuadamente hacia sus contrapartes exactos, los valores observados de una prueba a otra fueron muy diferentes; de hecho, en la prueba de Kruskal-Wallis no puede rechazarse la hipótesis nula de igualdad entre los vectores de parámetros multinomiales del tratamiento y el control al nivel de significación  $\alpha$  de 5%; hecho que sí ocurre si la decisión se realiza con base a cualquiera de las otras tres pruebas. Así, una vez más, se corrobora la opinión de que debe escogerse cuidadosamente y según la naturaleza de los datos qué prueba de significación aplicar.

### 6.1.2. Tablas de contingencia de dimensión $R \times C$ doblemente ordenadas

*Ejemplo 7:* La tabla de contingencia 19, de dimensión  $3 \times 3$  fue utilizada como ejemplo por Agresti y Wackerly, (1977); para ilustrar el comportamiento de las pruebas de significación en tablas de contingencia doblemente ordenadas.

**Tabla 19.** Tabla de contingencia doblemente ordenada utilizada por Agresti y Wackerly como ejemplo ilustrativo.

	Alto	Medio	Bajo	Total
Alto	6	4	2	12
Medio	4	4	4	12
Bajo	2	4	6	12
Total	12	12	12	36

Fuente: Agresti y Wackerly, 1977.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 20.

**Tabla 20.** Resultados de los  $p$ -valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 7.

	PRUEBAS DE SIGNIFICACIÓN	
	Jonckheere-Terpstra	Asociación Lineal-Lineal
$p$ -valor exacto	0,05282733	0,06329241
$p$ -valor asintótico	0,04871004	0,04860657
$\Delta$	0,00411729	0,01468584

Para la tabla de contingencia 19, la cual presenta totales de fila balanceados y tamaño de muestra pequeño,  $N = 36$ ; se observa en la tabla 20 que para ambas pruebas, la prueba de Jonckheere-Terpstra y la prueba de asociación lineal-lineal, las decisiones tomadas al nivel de significación  $\alpha$  de 5% hubieran sido diferentes según se utilizaran métodos exactos o asintóticos; así, de utilizar métodos exactos, la decisión hubiera sido la de no rechazar la hipótesis nula de cierto grado de asociación entre las categorías de filas y columnas; y de utilizar métodos asintóticos se hubiera concluido que existe cierto grado de asociación entre las categorías de filas y columnas.

Es interesante notar que si se tomara una decisión al nivel de significación  $\alpha$  de 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 de Jonckheere-Terpstra; en la cual, donde a pesar de que pueden utilizarse metodologías asintóticas, la decisión no coincide con la metodología exacta, hecho que no debería ocurrir por cuanto que  $\Delta$  es inferior al umbral  $\Delta^*$ ; ante esta situación Berger, (2000); advierte que al utilizar a  $\Delta$  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  $\Delta^* = 0,01$ .

Por otra parte, es importante señalar que para la prueba de asociación lineal-lineal la diferencia entre el  $p$ -valor exacto y el  $p$ -valor asintótico  $\Delta$  fue superior al umbral  $\Delta^*$ , lo que indica, siguiendo la recomendación de Berger, (2000); que para este ejemplo en particular, las decisiones se tomarían con base al  $p$ -valor exacto; en cambio para la prueba de Jonckheere-Terpstra ocurre exactamente lo contrario, ya que  $\Delta$  fue inferior al umbral  $\Delta^*$  y las decisiones se tomarían con base al  $p$ -valor asintótico; sin embargo, tales decisiones serían contrarias en ambas pruebas.

La convergencia más rápida mostrada por parte de la prueba de Jonckheere-Terpstra puede deberse a la naturaleza del estadístico de prueba utilizado, el cual es la suma de los  $r(r-1)/2$  posibles estadísticos de Wilcoxon-Mann-Whitney generados tomando de dos en dos las  $r$  filas de la tabla de contingencia; tal estadístico de prueba puede generar una distribución con un mayor número de puntos y por ende, menos discreta que la generada por la prueba de asociación lineal-lineal, el cual sólo contempla la suma de los productos de las marcas de filas y columnas, las cuales en este estudio son además, enteras.

*Ejemplo 8:* La siguiente tabla de contingencia reportada por Agresti, (1990); y citada por Lang, (1996); muestra el resultado de un estudio donde se clasificó el tipo de

operación y el grado de trastorno de vaciamiento gastrointestinal (*dumping*) en cierto hospital. El tipo de operación presentó cuatro categorías ordenadas: A: *dumping* y vagotomía; B: 25% de corte y vagotomía; C: 50% de corte y vagotomía; D: 75% de corte. La severidad del *dumping* fue clasificada en tres grados: N: ninguno; S: ligero, M: moderado. Los resultados se presentan en la tabla 21.

**Tabla 21.** Severidad de *dumping* y tipo de operación realizada.

SEVERIDAD DEL DUMPING				
OPERACIÓN	N	S	M	Total
A	8	6	3	17
B	12	4	4	20
C	11	6	2	19
D	7	7	4	18
<b>Total</b>	38	23	13	74

Fuente: Lang, 1996.

Los *p*-valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 22.

**Tabla 22.** Resultados de los *p*-valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 8.

PRUEBAS DE SIGNIFICACIÓN		
	Jonckheere-Terpstra	Asociación Lineal-Lineal
<i>p</i> -valor exacto	0,65210143	0,72966162
<i>p</i> -valor asintótico	0,64794457	0,69274691
$\Delta$	0,00415686	0,03691471

Para la tabla de contingencia 21, la cual presenta un tamaño de muestra moderado,  $N = 74$ ; y totales marginales de filas balanceados; se observa en la tabla 22 que para la prueba de Jonckheere-Terpstra la diferencia  $\Delta$  entre los  $p$ -valores exacto y asintótico fue inferior al umbral  $\Delta^*$  y que para la prueba de asociación lineal-lineal esta diferencia  $\Delta$  fue superior al umbral  $\Delta^*$ ; aunque las decisiones tomadas al nivel de significación  $\alpha$  de 5% hubieran sido iguales según estos resultados, en la prueba de Jonckheere-Terpstra pudiera usarse el  $p$ -valor asintótico y en la prueba de asociación lineal-lineal el  $p$ -valor exacto. Se observa además que para este tamaño de muestra, la prueba de Jonckheere-Terpstra parece converger mucho más rápido que la prueba de asociación lineal-lineal.

*Ejemplo 9:* El siguiente ejemplo, provisto por Yates, (1948); y citado por Patefield, (1982); muestra los resultados de un estudio realizado sobre 1019 niños en el cual se relacionan las condiciones bajo las cuales se realiza determinada tarea, ordenada en forma creciente, y la calificación adjudicada por el docente, ordenada en forma decreciente, los resultados se presentan en la tabla 23.

**Tabla 23.** Calificación del docente y condiciones en la que se realizó determinada tarea.

CALIFICACIÓN DEL DOCENTE				
CONDICIONES	A	B	C	Total
A	141	131	36	308
B	67	66	14	147
C	114	143	38	295
D	79	72	28	179
E	39	35	16	70
<b>Total</b>	440	447	132	1019

Fuente: Patefield, 1982.

Los  $p$ -valores exactos y asintóticos junto con los valores  $\Delta$  para las pruebas consideradas en este estudio se presentan en la tabla 24.

**Tabla 24.** Resultados de los  $p$ -valores exactos y asintóticos con el  $\Delta$  correspondiente para los datos del ejemplo 9.

PRUEBAS DE SIGNIFICACIÓN		
	Jonckheere-Terpstra	Asociación Lineal-Lineal
$p$ -valor exacto	0,16945000	0,13010000
$p$ -valor asintótico	0,16445076	0,12696289
$\Delta$	0,00499924	0,00313711

Para la tabla de contingencia 23, la cual presenta un tamaño de muestra grande,  $N = 1019$ ; y totales de filas moderadamente desbalanceados; se observa en la tabla 24 que tanto en la prueba de Jonckheere-Terpstra como en la prueba de

asociación lineal-lineal; los  $p$ -valores asintóticos convergen a sus contrapartes exactos ya que sus diferencias  $\Delta$  fueron menores que el umbral  $\Delta^*$ ; este fenómeno puede deberse simultáneamente tanto al tamaño de la muestra como a la dimensión de la tabla de contingencia, la cual es  $5 \times 3$ ; lo que genera un conjunto de referencia  $\Gamma$  de tablas de contingencia condicionadas muy grande y cuyas distribuciones de probabilidad para ambos estadísticos de prueba presentarán muchos puntos; es decir, serán poco discretas; esta situación se evidenció debido a el hecho de que se tuvieron que estimar los  $p$ -valores exactos en ambas pruebas utilizando el método de ensayos de Monte Carlo, ya que el computador utilizado no fue capaz de calcular el valor del  $p$ -valor exacto; por lo que es natural la convergencia de los  $p$ -valores asintóticos a los  $p$ -valores exactos.

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 ambas metodologías dependiendo de la situación considerada, esto induce a pensar que las conclusiones a las que se han llegado en muchos trabajos publicados pueden diferir notablemente a las obtenidas 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 previa, se pudo ilustrar que el comportamiento de las metodologías exactas y asintóticas en tablas de contingencia dimensión  $R \times C$  simple y doblemente ordenadas 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.
- Elección de las marcas de filas y columnas.

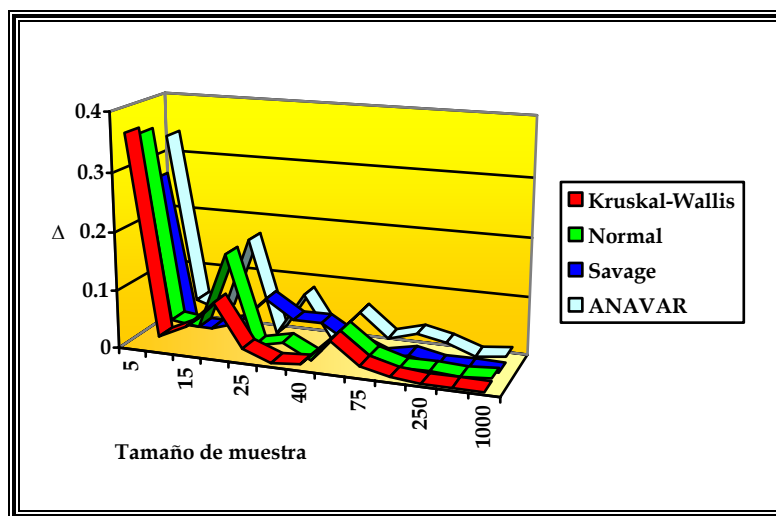
Por estas razones se realizó la simulación de tablas de contingencia simple y doblemente ordenadas con la finalidad de estudiar con más detalle este comportamiento y así recomendar el uso de una u otra metodología en función de los factores anteriores.

Siguiendo la metodología establecida con la simulación de las 1404 tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas y las 936 tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas, se calcularon los  $p$ -valores exactos, asintóticos y en aquellos casos en los cuales el tiempo de

procesamiento computacional fue superior a 15 minutos, los estimados de los  $p$ -valores exactos mediante ensayos de Monte Carlo. Debido a que estos  $p$ -valores presentaron resultados muy variables entre sí; resultó conveniente estudiar el módulo de la diferencia entre estos  $p$ -valores exactos y asintóticos, denotado por  $\Delta$ , para llevar a cabo su comparación, tal como se planteó en la sección de materiales y métodos. Estas diferencias  $\Delta$  originaron un total de 108 tablas de resultados con sus respectivos gráficos para las tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas y 72 tablas de resultados con sus respectivos gráficos para las tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas, presentes en forma digital en el CD correspondiente al anexo 2; tales gráficos muestran las tendencias entre las diferencias de los  $p$ -valores y constituyen una forma de apreciar la convergencia de los  $p$ -valores asintóticos a los exactos para cada una de las pruebas aplicadas y en cada situación considerada. Con fines ilustrativos, la figura 1 muestra uno de tales gráficos correspondiente a las pruebas de hipótesis aplicadas sobre tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas, ya que resultaría impráctico presentar todos los gráficos en esta sección.

En la figura 1 puede observarse que para este caso en particular todos los  $p$ -valores asintóticos convergieron a sus contrapartes exactos; es decir, sus diferencias  $\Delta$  fueron inferiores al umbral  $\Delta^*$  pero tal convergencia fue alcanzada para tamaños de muestra diferentes; para las pruebas de Kruskal-Wallis y marcas

normales esto se logró para  $N = 100$  ; en la prueba de marcas de Savage el comportamiento entre ambos  $p$ -valores fue equivalente a partir de  $N = 250$  y el ANAVAR con marcas arbitrarias alcanzó la convergencia a partir de  $N = 500$ , lo cual muestra que la elección de las marcas de las filas en la tabla de contingencia afecta a la tasa de convergencia entre los  $p$ -valores exactos y asintóticos.



*Figura 1. Valores de la medida de similitud  $\Delta$  según el tamaño de muestra para las pruebas consideradas en tablas de contingencia de dimensión  $3 \times 3$  simplemente ordenadas desbalanceadas y vectores de probabilidades multinomiales simétricos.*

A partir de cada una de estas tablas y sus respectivos gráficos, se determinaron los tamaños de muestras críticos; es decir, aquellos tamaños de muestra mínimos  $\{N_c\}$  para los cuales el módulo de la diferencia de los  $p$ -valores exactos y asintóticos hayan sido menores que el umbral establecido  $\Delta^* = 0,01$ .

Tales tamaños de muestra críticos permiten observar la tendencia de la convergencia de los  $p$ -valores exactos a los asintóticos según los diferentes esquemas poblacionales mostrados. Éstos se presentan en las tablas 25, 26, 27 y 28 para las pruebas aplicadas en tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas al final de la sección 6.2.1 y en las tablas 29 y 30 para las pruebas aplicadas en tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas al final de la sección 6.2.2.

### *6.2.1. Simulación en tablas de contingencia de dimensión $R \times C$ simplemente ordenadas*

En el caso de las pruebas aplicadas en tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas, en las tablas 25, 26, 27 y 28 la tendencia observada es que el tamaño de muestra crítico observado disminuye a medida que la dimensión de la tabla de contingencia aumenta, independientemente de la prueba aplicada o el vector de parámetros multinomiales considerado; adicionalmente, se observó que cuando la hipótesis nula es falsa o la población es uniforme, los tamaños de muestra críticos  $\{N_c\}$  tienden a ser menores que los observados en las poblaciones simétrica o decreciente; este comportamiento pudiera explicarse con base a dos posibles razones:

- (i) En el caso de poblaciones con esquema de vectores de parámetros multinomiales uniformes, la distribución de probabilidad de los estadísticos de las pruebas aplicadas presentan más puntos que las correspondientes a las estructuras poblacionales simétrica y decreciente; debido a que el conjunto de referencia  $\Gamma$  de tablas de contingencia condicionadas es por definición mucho más grande en las poblaciones uniformes, lo cual permite la convergencia más rápida de los  $p$ -valores asintóticos a los exactos en tales poblaciones.
- (ii) Cuando la hipótesis nula es falsa, las tablas de contingencia consideradas para el cálculo del  $p$ -valor exacto pertenecientes al conjunto de referencia reducido  $\Gamma^*$  tenderán a presentar medidas de discrepancia  $D(\mathbf{y})$  cada vez más cercanas al valor crítico  $d_\alpha(\mathbf{y}) = \min\{d(\mathbf{y}) : \Pr[D(\mathbf{y}) \geq d(\mathbf{y})] \leq \alpha\}$  y por ende, los  $p$ -valores correspondientes a las tablas de contingencia observadas tenderán a presentar valores cercanos al nivel de significación  $\alpha$ ; y en la medida que el estadístico de prueba observado  $D(\mathbf{x})$  presente valores cada vez mayores, los  $p$ -valores exactos y asintóticos tenderán a ser pequeños y a su vez menores que el valor del nivel de significación  $\alpha$ ; bajo tal escenario, puede esperarse que las diferencias  $\Delta$  entre ambos  $p$ -valores será bastante

pequeña y que se superará el umbral  $\Delta^*$  para tamaños de muestra relativamente menores que en las otras configuraciones poblacionales.

Las prueba de Kruskal-Wallis, marcas normales y marcas de Savage mostraron un comportamiento similar, mientras que el ANAVAR con marcas arbitrarias mostró un comportamiento diferente a las anteriores, ya que en esta última prueba se presentaron dificultades para encontrar convergencia entre ambos  $p$ -valores, ratificando así los argumentos expuestos por Graubard y Korn, (1987); Permutt y Berger, (2000); e Ivanova y Berger, (2001); en que se debe ser muy prudente en la elección de marcas y que por lo general, las marcas enteras tienden a originar estadísticos con distribuciones de probabilidad más discretas que las de aquellos estadísticos cuyas marcas estén basadas en números reales y racionales. Esto pudiera explicarse observando la naturaleza del estadístico de prueba

$$D(\mathbf{y}) = (S_w)^{-2} \sum_{i=1}^r [T_i - E(T_i)]^2 / m_{i+}$$

el cual depende de las cantidades

$$T_i = \sum_{j=1}^c w_j y_{ij}$$

donde se observa que si las marcas  $w_j$  no son enteras, el estadístico presentará una distribución menos discreta que en el caso de que sí lo fueran, lo que permite la convergencia mucho más rápida en el caso de las pruebas de Kruskal-Wallis,

marcas normales y marcas de Savage y más lenta en el caso del ANAVAR con marcas arbitrarias. Por otra parte, como en este estudio el ANAVAR presenta marcas lineales y enteras, la distribución del estadístico tenderá a presentar muchos menos puntos que en las otras tres pruebas aplicadas, lo que pudiera originar que en esta prueba, los  $p$ -valores converjan más lentamente que en los otros casos.

En general, para todas las pruebas y dimensiones consideradas en las tablas simplemente ordenadas, los tamaños de muestra críticos tienden a ser menores en las tablas de contingencia balanceadas más que en las desbalanceadas, ver tablas 25, 26, 27 y 28; lo que pudiera deberse a que las tablas balanceadas presentan un conjunto de referencia  $\Gamma$  más grande que en caso de las tablas desbalanceadas, y esta característica origina distribuciones menos discretas para los estadísticos de prueba en el caso de las tablas balanceadas, lo que hace que la convergencia de los  $p$ -valores exactos y asintóticos sea más rápida en el caso de las tablas balanceadas que en las desbalanceadas.

*Tabla 25. Tamaños de muestra críticos observados para el umbral  $\Delta^* = 0,01$  en la prueba de Kruskal-Wallis para tablas de contingencia simplemente ordenadas.*

Dimensión	Balanceadas				Desbalanceadas			
	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
2×3	> 1000	250	500	1000	1000	> 1000	500	250
2×4	40	40	50	250	100	250	75	75
2×5	50	50	75	40	50	100	75	100
3×3	75	75	100	75	250	100	100	50
3×4	25	40	40	30	75	75	100	40
3×5	30	40	50	40	75	75	100	75
4×3	25	250	30	50	75	50	40	250
4×4	50	30	50	30	75	75	100	75
4×5	30	75	40	40	50	100	100	75
5×3	40	75	75	40	100	75	250	100
5×4	20	40	30	40	50	75	250	50
5×5	40	75	40	30	100	100	100	50



*Tabla 26. Tamaños de muestra críticos observados para el umbral  $\Delta^* = 0,01$  en la prueba de marcas normales para tablas de contingencia simplemente ordenadas.*

Dimensión	Balanceadas				Desbalanceadas			
	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
2×3	> 1000	250	500	1000	1000	> 1000	500	> 1000
2×4	75	250	100	75	75	100	250	100
2×5	50	75	40	100	75	50	250	40
3×3	250	75	100	75	250	100	250	50
3×4	40	75	40	40	75	75	100	40
3×5	25	40	25	50	75	30	40	50
4×3	75	250	25	50	75	50	40	250
4×4	50	25	40	30	75	75	75	75
4×5	50	75	40	30	50	75	100	75
5×3	30	75	75	40	100	75	250	75
5×4	30	40	40	40	50	75	250	50
5×5	75	75	40	30	250	100	100	250

*Tabla 27. Tamaños de muestra críticos observados para el umbral  $\Delta^* = 0,01$  en la prueba de marcas de Savage para tablas de contingencia simplemente ordenadas.*

Dimensión	Balanceadas				Desbalanceadas			
	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
2x3	250	100	500	250	100	250	500	100
2x4	40	75	50	40	100	75	500	75
2x5	40	40	50	40	40	100	100	40
3x3	75	75	50	40	250	250	100	100
3x4	30	50	50	40	75	75	75	40
3x5	10	40	75	75	75	75	250	100
4x3	40	250	40	40	50	50	250	250
4x4	50	50	100	30	250	75	250	75
4x5	50	50	75	50	75	250	75	250
5x3	40	75	75	30	100	250	100	100
5x4	40	75	100	50	50	40	25	75
5x5	75	40	250	75	100	30	250	250

*Tabla 28. Tamaños de muestra críticos observados para el umbral  $\Delta^* = 0,01$  en el análisis de varianza de una vía con marcas arbitrarias para tablas de contingencia simplemente ordenadas.*

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

### *6.2.2. Simulación en tablas de contingencia de dimensión $R \times C$ doblemente ordenadas*

En relación a las tablas de contingencia doblemente ordenadas, para el caso de la prueba de Jonckheere-Terpstra, el comportamiento fue similar tanto en las tablas de contingencia balanceadas como en las desbalanceadas e incluso si la hipótesis nula es cierta o falsa, ver tablas 29 y 30; este comportamiento homogéneo pudiera deberse a la naturaleza misma de la prueba, puesto que la medida de discrepancia en esta prueba corresponde a la suma de los  $r(r-1)/2$  posibles estadísticos de Wilcoxon-Mann-Whitney, tomados a partir de las  $r$  categorías de filas de la tabla de contingencia. Por supuesto, tal estadístico presentará una distribución de probabilidad con muchos puntos, lo que conduce a que la convergencia del  $p$ -valor asintótico al exacto se alcanza de manera muy rápida. Cabe destacar que esta prueba fue la que tomó más tiempo de procesamiento computacional en relación a todas las demás.

En el caso de la prueba de asociación lineal ocurrió algo similar, pero mucho más marcado que en el ANAVAR con marcas arbitrarias; la convergencia resultó errática para todas las estructuras poblacionales, excepto cuando la hipótesis nula es falsa, ver tablas 29 y 30; este comportamiento errático pudiera deberse a que se simuló una sola tabla para cada escenario considerado.

Los tamaños de muestra críticos, en general resultaron mayores en comparación con la prueba de Jonckheere-Terpstra, lo que evidencia la falta de convergencia de los  $p$ -valores asintóticos a los exactos en la prueba de asociación lineal-lineal.

Pareciera que la conjetura y la recomendación de Permutt y Berger, (2000); e Ivanova y Berger, (2001); referente a la elección de marcas no enteras también es válida en tablas de contingencia doblemente ordenadas, por cuanto que la distribución del estadístico de la prueba de asociación lineal-lineal, el cual presenta marcas enteras en las categorías de filas y columnas, presenta mucho menos puntos que la correspondiente a la prueba de Jonckheere-Terpstra, lo cual hace que la convergencia sea más lenta en el primer caso.

*Tabla 29. Tamaños de muestra críticos observados para el umbral  $\Delta^* = 0,01$  en la prueba de Jonckheere-Terpstra para tablas de contingencia doblemente ordenadas.*

Dimensión	Balanceadas				Desbalanceadas			
	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa	H <sub>0</sub> Cierta			H <sub>0</sub> Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
3×3	100	40	50	75	250	75	40	50
3×4	75	30	40	15	30	75	40	30
3×5	40	30	25	30	40	40	40	50
4×3	40	250	40	25	30	40	40	30
4×4	30	30	40	25	40	30	40	40
4×5	30	30	40	40	30	20	30	30
5×3	25	250	30	30	40	40	30	40
5×4	40	40	40	25	30	40	20	30
5×5	25	25	40	30	40	30	40	40

*Tabla 30. Tamaños de muestra críticos observados para el umbral  $\Delta^* = 0,01$  en la prueba de asociación lineal-lineal para tablas de contingencia doblemente ordenadas.*

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

### 6.2.3. *Tamaños de muestra minimax*

A partir de todos los resultados anteriores, se tomó una decisión minimax para recomendar el uso de una metodología u otra en función del tamaño de la muestra y según la dimensión de la tabla de contingencia con su correspondiente balanceo y para cada una de las pruebas. Tales resultados se presentan en la tabla 31 para las pruebas aplicadas en tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas y en la tabla 32 para las pruebas aplicadas en tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas.

A partir de los resultados mostrados en las tablas 31 y 32; se evidencia de forma mucho más clara que la tasa de convergencia es más lenta en las tablas desbalanceadas en relación a las tablas balanceadas tanto para las tablas de contingencia simplemente ordenadas como para las tablas de contingencia doblemente ordenadas; y como se dijo anteriormente, el tamaño de muestra crítico tiende a disminuir con la dimensión de la tabla de contingencia; estos resultados ofrecen una sugerencia con referencia a partir de cuál tamaño de muestra se pueden utilizar pruebas asintóticas en vez de pruebas exactas para el umbral  $\Delta^*$  considerado en este trabajo.

Particularmente, en el caso de la prueba de Jonckheere-Terpstra para las tablas de contingencia doblemente ordenadas, la convergencia observada de los  $p$ -valores asintóticos a los exactos fue muy rápida, lo que permite utilizar



metodologías asintóticas para tamaños de muestra menores en comparación con otras pruebas consideradas en este estudio.

En el caso de la prueba de asociación lineal-lineal, considerando las marcas utilizadas en este estudio, pudiera ser prudente utilizar solamente pruebas exactas ya que los tamaños de muestra minimax resultaron ser muy grandes.

Si bien este estudio es preliminar y exploratorio y está basado en una simulación limitada, debido principalmente a lo exigente del tiempo de procesamiento computacional por parte de las pruebas exactas, los resultados obtenidos permiten establecer un punto de partida para realizar estudios más detallados en cuanto a la convergencia de los métodos asintóticos a los métodos exactos en tablas de contingencia ordenadas.

*Tabla 31. Tamaños de muestra minimax observados con el umbral  $\Delta^* = 0,01$  para las pruebas consideradas en tablas de contingencia de dimensión  $R \times C$  simplemente ordenadas.*

Dimensión	Kruskal-Wallis		Marcas normales		Marcas de Savage		ANAVAR	
	$N_{\text{minimax}}$ tablas balanceadas	$N_{\text{minimax}}$ tablas desbalanceadas	$N_{\text{minimax}}$ tablas balanceadas	$N_{\text{minimax}}$ tablas desbalanceadas	$N_{\text{minimax}}$ tablas balanceadas	$N_{\text{minimax}}$ tablas desbalanceadas	$N_{\text{minimax}}$ tablas balanceadas	$N_{\text{minimax}}$ tablas desbalanceadas
2×3	> 1000	> 1000	> 1000	> 1000	500	500	> 1000	> 1000
2×4	250	250	250	250	75	500	> 1000	> 1000
2×5	75	100	100	250	50	100	> 1000	> 1000
3×3	100	250	250	250	75	250	500	500
3×4	40	100	75	100	50	75	100	500
3×5	50	100	50	75	75	250	250	100
4×3	250	250	250	250	250	250	250	250
4×4	50	100	50	75	100	250	100	250
4×5	75	100	75	100	75	250	250	250
5×3	75	250	75	250	75	250	250	250
5×4	40	250	40	250	100	75	100	75
5×5	75	100	75	250	250	250	250	250

*Tabla 32. Tamaños de muestra minimax observados con el umbral  $\Delta^* = 0,01$  para las pruebas consideradas en tablas de contingencia de dimensión  $R \times C$  doblemente ordenadas.*

Dimensión	Prueba de Jonckheere-Terpstra		Prueba de asociación lineal-lineal	
	$N_{\text{minimax}}$ tablas balanceadas	$N_{\text{minimax}}$ tablas desbalanceadas	$N_{\text{minimax}}$ tablas balanceadas	$N_{\text{minimax}}$ tablas desbalanceadas
3×3	100	250	> 1000	1000
3×4	75	75	> 1000	1000
3×5	40	50	> 1000	1000
4×3	250	40	1000	> 1000
4×4	40	40	> 1000	1000
4×5	40	30	1000	> 1000
5×3	250	40	> 1000	1000
5×4	40	40	> 1000	> 1000
5×5	40	40	1000	250

## 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, prueba aplicada y de las marcas utilizadas, tal como se evidenció en los ejemplos encontrados en la literatura internacional.
- El tamaño de muestra crítico tendió a disminuir con la dimensión de la tabla de contingencia independientemente de la prueba y de la estructura poblacional utilizada, excepto en la prueba de asociación lineal-lineal para tablas de contingencia doblemente ordenadas.
- El tamaño de muestra crítico en las tablas de contingencia simplemente ordenadas tendió a ser mayor en el ANAVAR con marcas arbitrarias debido posiblemente al uso de marcas enteras.
- En el caso de la prueba de asociación lineal-lineal para tablas de contingencia doblemente ordenadas, el tamaño de muestra crítico tendió a ser mayor en comparación a la prueba de Jonckheere-Terpstra.
- En general, para todos los casos estudiados se encontró que los  $p$ -valores exactos y asintóticos presentaron diferencias notables para tamaños de muestras pequeños; sin embargo, empezaron a converger a partir de cierto

tamaño de muestra crítico, variando este último según la dimensión, prueba o balanceo de la tabla.

- La convergencia observada fue más rápida en tablas de contingencia balanceadas que en tablas de contingencia desbalanceadas, excepto en las pruebas para tablas de contingencia doblemente ordenadas, donde la convergencia observada fue similar en ambos casos e incluso, ligeramente más rápida en tablas de contingencia desbalanceadas.
- Si bien este es un estudio preliminar, puede recomendarse el uso de métodos asintóticos a partir de los tamaños de muestra minimax obtenidos anteriormente y usar métodos exactos para tamaños de muestra menores.
- Realizar estudios de simulación con un número mayor de tablas de contingencia simuladas al considerado en este estudio.
- Considerar el estudio del comportamiento de los  $p$ -valores exacto y asintótico para diferentes valores del umbral  $\Delta^*$ .
- Considerar esquemas de desbalanceo mucho más marcados para las tablas de contingencia, ya que esto pudiera afectar considerablemente la tasa de convergencia de los  $p$ -valores asintóticos a los exactos.
- Considerar otros esquemas de marcas arbitrarias diferentes de las lineales para el ANAVAR con marcas arbitrarias en tablas de contingencia

simplemente ordenadas, y en la prueba de asociación lineal-lineal en tablas de contingencia doblemente ordenadas.

- Llevar a cabo estudios de potencia para las pruebas aplicadas en tablas de contingencia considerando marcas arbitrarias diferentes a las de Wilcoxon-Mann-Whitney, Normales y de Savage.
- Mientras la capacidad de procesamiento del computador lo permita, pueden usarse pruebas exactas, o en su defecto, puede obtenerse un estimado del  $p$ -valor exacto mediante el método de simulación de Monte Carlo.
- En tablas de contingencia doblemente ordenadas, es preferible aplicar la prueba de Jonckheere-Terpstra, dado que la convergencia de los  $p$ -valores asintóticos a los exactos fue más rápida en esta prueba que la prueba de asociación lineal-lineal.
- 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. APÉNDICE

8.1. Deducción de la distribución de probabilidad exacta para pruebas de permutación condicionadas en tablas de contingencia de dimensión  $R \times C$  según el esquema de muestreo de producto de poblaciones multinomiales independientes

Dada la siguiente tabla de contingencia  $R \times C$  de la tabla 33:

*Tabla 33. Tabla de contingencia de dimensión  $R \times C$ .*

		Columnas				Total
		1	2	...	$c$	
Filas	1	$y_{11}$	$y_{12}$	...	$y_{1c}$	$m_{1+}$
	2	$y_{21}$	$y_{22}$	...	$y_{2c}$	$m_{2+}$
	⋮	⋮	⋮	⋮	⋮	⋮
	$r$	$y_{r1}$	$y_{r2}$	...	$y_{rc}$	$m_{r+}$
Total		$n_{+1}$	$n_{+2}$	...	$n_{+c}$	$N$

La distribución de probabilidad de algún arreglo en una tabla de contingencia  $R \times C$  de los  $\{y_{ij}\}$  puede considerarse multinomial con parámetros  $N$  y  $\{\pi_{ij}\}$ ,  $i = 1, \dots, r$ ,  $j = 1, \dots, c$ , dado que los totales de las filas se consideran fijos, pero no el total de las columnas, es decir, tal distribución queda:

$$\frac{N!}{y_{11}! y_{12}! \dots y_{rc}!} \pi_{11}^{y_{11}} \pi_{12}^{y_{12}} \dots \pi_{rc}^{y_{rc}} = \frac{N!}{\prod_i \prod_j y_{ij}!} \prod_i \prod_j \pi_{ij}^{y_{ij}}$$

Esta distribución se ilustra en el caso de una tabla de contingencia de dimensión  $2 \times 2$  en la tabla 34.

*Tabla 34. Tabla de contingencia observada de dimensión  $2 \times 2$ .*

		Columnas		
		1	2	Total
Filas	1	7	12	19
	2	5	1	6
Total		12	13	25

La probabilidad de tal arreglo es:

$$\frac{25!}{7!12!5!1!} \pi_{11}^7 \pi_{12}^{12} \pi_{21}^5 \pi_{22}^1$$

En muchas situaciones, en cambio, para la obtención de la distribución de algún arreglo en una tabla de contingencia  $R \times C$  de las frecuencias conjuntas  $\{y_{ij}\}$ , se considera que la variable respuesta, representada por las columnas, depende de  $r$  variables multinomiales independientes representadas por las filas, es decir, la distribución de la variable respuesta  $j$  está condicionada al hecho de que la variable causa toma el valor  $i$ . Considerando los totales de fila  $m_{i+}$  fijos, la distribución de la variable respuesta para cada valor  $i$  de la variable causa es multinomial con parámetros  $N$  y  $\{\pi_{ji}\}$ ,  $i = 1, \dots, r$ ,  $j = 1, \dots, c$ , es decir,



$$\frac{m_{i+}!}{\prod_j y_{ij}!} \prod_j \pi_{j|i}^{y_{ij}} ; \text{ con } \sum_j y_{ij} = m_{i+}$$

Para la derivación de la distribución exacta condicional de un arreglo de las  $\{y_{ij}\}$  en alguna tabla de contingencia, se debe considerar:

- Los totales de filas  $\{m_{i+}\}$  se consideran fijos.
- Se suponen que las filas están constituidas por  $r$  variables multinomiales independientes con parámetros  $m_{i+}$  y  $\{\pi_{j|i}\}$ .
- Bajo la hipótesis nula de independencia, se cumple:

$$\pi_{j|1} = \pi_{j|2} = \dots = \pi_{j|r} = \pi_{+j}$$

Ya que, bajo independencia se cumple:

$$\begin{aligned} \pi_{j|1} &= \pi_{1j} / \pi_{1+} = \pi_{1+} \pi_{+j} / \pi_{1+} = \pi_{+j} \\ \pi_{j|2} &= \pi_{2j} / \pi_{2+} = \pi_{2+} \pi_{+j} / \pi_{2+} = \pi_{+j} \\ &\vdots \\ \pi_{j|r} &= \pi_{rj} / \pi_{r+} = \pi_{r+} \pi_{+j} / \pi_{r+} = \pi_{+j} \end{aligned}$$

Como se supone que las  $r$  filas son multinomiales independientes con parámetros  $m_{i+}$  y  $\{\pi_{j|i}\}$ , cada una de ellas tiene distribución

$$\frac{m_{i+}!}{\prod_j y_{ij}!} \prod_j \pi_{j|i}^{y_{ij}} ; i = 1, \dots, r$$

Y la probabilidad conjunta para las  $r$  filas viene dada por:

$$\left( \frac{m_{1+}!}{\prod_j y_{1j}!} \prod_j \pi_{j|1}^{y_{1j}} \right) \left( \frac{m_{2+}!}{\prod_j y_{2j}!} \prod_j \pi_{j|2}^{y_{2j}} \right) \cdots \left( \frac{m_{r+}!}{\prod_j y_{rj}!} \prod_j \pi_{j|r}^{y_{rj}} \right)$$

Lo cual queda

$$\prod_i \left( \frac{m_{i+}!}{\prod_j y_{ij}!} \prod_j \pi_{j|i}^{y_{ij}} \right)$$

El cual, bajo hipótesis nula de independencia se puede escribir como:

$$\prod_i \left( \frac{m_{i+}!}{\prod_j y_{ij}!} \prod_j \pi_{+j}^{y_{ij}} \right) = \frac{\prod_i m_{i+}!}{\prod_i \prod_j y_{ij}!} \prod_i \prod_j \pi_{+j}^{y_{ij}} = \frac{\prod_i m_{i+}!}{\prod_i \prod_j y_{ij}!} \prod_j \pi_{+j}^{\sum_i y_{ij}} = \frac{\prod_i m_{i+}!}{\prod_i \prod_j y_{ij}!} \prod_j \pi_{+j}^{n_{+j}}$$

Se observa pues, que la distribución multinomial conjunta del producto de las  $r$  filas depende de  $\{\pi_{+j}\}$  y  $\{n_{+j}\}$ . Para eliminar los parámetros desconocidos  $\{\pi_{+j}\}$  se condicionan los totales de las columnas que presenten tamaño fijo, entonces, los totales de columnas  $\{n_{+j}\}$  tendrán distribución multinomial con parámetros  $\{\pi_{+j}\}$  y  $\{n_{+j}\}$ ; es decir:

$$\frac{N!}{\prod_j n_{+j}!} \prod_j \pi_{+j}^{n_{+j}}$$

La distribución de probabilidad conjunta de  $\{y_{ij}\}$  y de  $\{n_{+j}\}$  es idéntica la distribución de los  $\{y_{ij}\}$ , dado que los  $\{y_{ij}\}$  determinan a los  $\{n_{+j}\}$ ; así la distribución condicional dado los totales de filas y columnas fijos, es el cociente de ambas distribuciones, es decir:

$$\frac{\prod_i m_{i+}!}{\prod_i \prod_j y_{ij}!} \prod_j \pi_{+j}^{n_{+j}} = \frac{\prod_i m_{i+}! \prod_j n_{+j}!}{N! \prod_i \prod_j y_{ij}!} \frac{N!}{\prod_j n_{+j}!} \prod_j \pi_{+j}^{n_{+j}}$$

La última distribución no depende de los parámetros desconocidos  $\{\pi_{+j}\}$  y permite calcular probabilidades exactas, (Agresti, 2002).

## 9. REFERENCIAS BIBIOGRÁFICAS

- AGRESTI, A. 2002. Categorical data analysis. Second edition. Wiley-Interscience. New Jersey. USA. 718 p.
- AGRESTI, A.; LANG, J.; MEHTA, C. 1993. Some empirical comparisons of exact, modified exact, and higher-order asymptotic tests of independence for ordered categorical variables. Communications in statistics. Simulation. Vol. 22, N° 1. pp. 1-18.
- 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.
- BAJORSKI, P.; PETKAU, J. 1999. Nonparametric two-sample comparisons of changes on ordinal responses. Journal of the American Statistical Association. Vol. 94, N° 447. September 1999. pp. 970-978.

- 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.
- 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.
- BAYLY, P.; MATTHEWS, J.; DOBSON, P.; PRICE, M.; THOMAS, D. 2001. In-hospital mortality from abdominal aortic surgery in Great Britain and Ireland: Vascular Anaesthesia Society audit. British Journal of Surgery. Vol. 88, N° 5. pp. 687-692.
- 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 *p*-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.
- CONOVER, W. 1971. Practical nonparametric statistics. John Wiley & Sons INC. New York. USA. 462 p.

CYTEL SOFTWARE. 2001. StatXact 5<sup>®</sup>. Statistical software for exact nonparametric inference. User manual. Cytel Software. New York. USA. 1123 p.

CYTEL SOFTWARE. 2003. StatXact 6<sup>®</sup> with Cytel studio. Statistical software for exact nonparametric inference. User manual. Cytel Software. New York. USA. 1398 p.

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

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

EDUARDES, M. 1997. Univariate random cut-points theory for the análisis of ordered categorical data. Journal of the American Statistical Association. Vol. 92, N<sup>o</sup> 439. September 1997. pp. 1114-1123.

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<sup>o</sup> 360. December 1977. pp. 859-862.

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

GRAUBARD, B.; KORN, E. 1987. Choice of column scores for testing independence in ordered  $2 \times K$  contingency tables. Biometrics. N° 43. June 1987. pp. 471-476.

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

IVANOVA, A.; BERGER, V. 2001. Drawbacks to integer scoring for ordered categorical data. Biometrics. Vol. 57. June 2001. pp. 567-570.

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.

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.

- LANG, J. 1996. On the partitioning of goodness-of-fit statistics for multivariate categorical response models. *Journal of the American Statistical Association*. Vol. 91, N° 435. September 1996. pp. 1017-1023.
- 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, 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  $2 \times 2$  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  $2 \times 2$  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  $2 \times 2$  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.; 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.

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.

- NAIR, V. 1987. Chi-squared-type tests for ordered alternatives in contingency tables. *Journal of the American Statistical Association*. Vol. 82, N° 397. March 1987. pp. 283-291.
- PATEFIELD, W. 1982. Exact tests for trends in ordered contingency tables. *Applied Statistics*. Vol. 31, N° 1. pp. 32-43.
- PERMUTT, T.; BERGER, V. 2000. A new look at ranks tests in ordered  $2 \times k$  contingency tables. *Communications in Statistics. Theory and Methods*. Vol. 29 (5&6). pp. 989-1003.
- 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.
- SIEGEL, S.; CASTELLAN, N. 1988. *Nonparametric statistics for the behavioral sciences*. Second edition. McGraw-Hill. New York. USA.

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.

## **10. ANEXOS**

*Anexo 1. 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 = 25
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_25.res m
quit
```

**Anexo 2.** *Valores de la medida de similitud  $\Delta$  según el tamaño de muestra para las pruebas consideradas en tablas de contingencia de dimensión  $R \times C$  simple y doblemente ordenadas.*