Saber, Universidad de Oriente, Venezuela. Vol. 28 Nº 1: 116-124. (2016)
ISSN: 2343-6468 Digital / Depósito Legal ppi 198702SU4231 ISSN: 1315-0162 Impreso / Depósito Legal pp 198702SU187

COMPARACIÓN DE LAS PRUEBAS DE PERMUTACIÓN Y ASINTÓTICAS APLICADAS EN TABLAS DE CONTINGENCIA DE DIMENSIÓN 2×2

COMPARISON BETWEEN PERMUTATION AND ASYMPTOTIC TESTS APPLIED TO 2×2 DIMENSION CONTINGENCY TABLES

LOURDES SÁNCHEZ-AGUILERA[1], LUIS PÉREZ-YBARRA[2], HARÚ MARTÍNEZ DE CORDERO[3], WILLIAM QUINTANA-RIVERO[4]

Universidad Central de Venezuela, [1]Núcleo Cagua, Facultad de Ingeniería, Ciclo Básico, Departamento de Matemáticas, Cagua, Venezuela, [3]Facultad de Agronomía, Escuela de Agronomía, Departamento de Ingeniería Agrícola, Maracay, Venezuela. Universidad de Carabobo, Sede Aragua, [2]Facultad de Ciencias de la Salud, Escuela de Bioanálisis, Departamento de Ciencias Básicas, Maracay, Venezuela, [4]Facultad de Ciencias Económicas y Sociales, Escuela de Administración Comercial y Contaduría Pública, Departamento de Matemática, Estadística y Técnicas Cuantitativas, Maracay, Venezuela. E-mail: Lmpy2005@gmail.com

RESUMEN

Se realizó un estudio comparativo entre pruebas de permutación y asintóticas, aplicadas a tablas de contingencia de dimensión 2×2, utilizando como medida de comparación la diferencia entre el *p*-valor exacto y su equivalente asintótico. Se analizaron cinco ejemplos publicados en la literatura científica internacional con el objeto de mostrar bajo cuáles condiciones ambos enfoques presentan un comportamiento diferente o un comportamiento equivalente para las pruebas de independencia de $\chi 2$, razón de verosimilitud y prueba exacta de Fisher. Los resultados mostraron que el comportamiento de las metodologías exacta y asintótica depende del tamaño de muestra, del balanceo y la dispersión de la tabla de contingencia, así como la prueba de independencia aplicada. Para los casos estudiados, se encontró que los *p*-valores exactos y asintóticos presentaron diferencias significativas para tamaños de muestras pequeños, sobre todo en tablas de contingencia desbalanceadas y dispersas, y tendieron a mostrar convergencia en la medida que el tamaño de muestra era mayor.

PALABRAS CLAVE: *p*-valor, tablas de contingencia balanceadas, tablas de contingencia desbalanceadas, tablas de contingencia dispersas, pruebas de independencia.

ABSTRACT

A comparative study was conducted between exact and asymptotic tests applied to 2×2 dimension contingency tables using the difference between the exact *p*-value and the asymptotic counterpart as a comparison measure. Five examples published in the international scientific literature were analyzed in order to show under what conditions the two approaches present a different or equivalent behavior for independence tests χ 2, likelihood ratio, and Fisher exact test. The results showed that the behavior of exact and asymptotic methods depends on the sample size, the balance and the dispersion of the contingency table and the independency test applied. For the studied cases, it was found that exact and asymptotic *p*-values show significant differences for small sample sizes, especially in unbalanced and sparse contingency tables, and had a trend to convergence as the sample size increased.

KEY WORDS: *p*-value, balanced contingency tables, unbalanced contingency tables, sparse contingency tables, independence tests.

INTRODUCCIÓN

Una tabla de contingencia es una tabla de distribución de frecuencias absolutas conjuntas muestrales de dos variables aleatorias clasificadas en categorías (Agresti 2002, 2007).

El análisis de tablas de contingencia presenta un amplio rango de aplicación, ya que permiten verificar si dos variables cualitativas o discretas, medidas en términos de frecuencias absolutas y agrupadas en un arreglo rectangular de filas y columnas, están o no asociadas (Agresti 2002, 2007, Daniel 2008). Existen métodos que permiten analizar tales resultados basados en algún estadístico cuya distribución asintótica es conocida y se encuentra tabulada, como por ejemplo, la prueba de independencia basada en la distribución χ^2 de Pearson (1900). Tales métodos se comportan adecuadamente cuando son aplicados a muestras grandes y balanceadas. Sin embargo, esto no necesariamente ocurre si el tamaño de muestra es pequeño, o si el tamaño de muestra es grande en tablas desbalanceadas y/o dispersas (Agresti 2002,

Recibido: junio 2015. Aprobado: octubre 2015
Versión final: febrero 2016

2007, Daniel 2008, Cytel Software 2010).

Una alternativa para dar solución a estos problemas la constituyen las pruebas de permutación, las cuales generalmente utilizan el mismo estadístico que las correspondientes aproximaciones asintóticas, pero basan sus inferencias en la distribución de las permutaciones de los datos en vez de la distribución teórica correspondiente. Esto permite a dichas pruebas presentar niveles de significación exactos, dado que los *p*-valores derivados de las mismas se calculan a partir de una distribución discreta y libre de parámetros desconocidos generada por remuestreo de los datos originales (Good 2000, Agresti 2002).

Las pruebas de permutación exigen alta capacidad de procesamiento computacional, por lo cual inicialmente estuvieron limitadas a estudios con tamaños de muestra pequeños, pero los avances en el área de la informática han permitido que estas pruebas se hayan comenzado a aplicar en el campo de la investigación en diferentes disciplinas con tamaños de muestras cada vez mayores (Good 2000, Agresti 2002, Fortin *et al.* 2002).

Si bien es cierto que en Venezuela se han publicado recientemente trabajos similares al presente, estos fueron realizados sobre tablas de contingencia de mayores dimensiones:

$$\left(R\times C\right),\ R\geq 2\wedge C\geq 2;\ R\times C\neq 2\times 2,$$

Tanto no ordenadas como simplemente ordenadas, donde R y C corresponden, respectivamente, al número de filas y columnas de las tablas de contingencia (Pérez *et al.* 2013, Pérez-Ybarra *et al.* 2014), no contemplando el caso particular de dimensión 2×2. Por esa razón, en este trabajo se comparó el comportamiento de los enfoques asintótico y exacto, mediante el cálculo, comparación y análisis de los *p*-valores generados por ambas metodologías, en pruebas estadísticas de independencia aplicadas a tablas de contingencia de dimensión 2×2, publicadas en la literatura científica internacional, considerando varios tamaños de muestra, dimensión y balanceo de las tablas, con la finalidad de mostrar de forma empírica el comportamiento diferencial de ambos enfoques, aplicados en ejemplos provenientes de diversas disciplinas.

MATERIALES Y MÉTODOS

Se escogieron cinco ejemplos publicados en la literatura científica internacional correspondientes a tablas de contingencia de dimensión 2×2. Estas tablas fueron escogidas de tal forma que abarcaran ejemplos provenientes de diferentes disciplinas, tales como medicina, educación y ciencias económicas, y al mismo tiempo presentaran diferentes escenarios para los tamaños de muestra (pequeños y grandes), balanceo (balanceadas y desbalanceadas) y dispersión de las tablas de contingencia. Se entiende por tablas de balanceadas aquellas con tamaños de filas o columnas similares, y tablas dispersas a aquellas que presenten frecuencias muy bajas o nulas (Good 2000, Cytel Software 2010).

Con respecto las tablas seleccionadas para el desarrollo de los ejemplos, se tiene lo siguiente: La tabla analizada en el primer ejemplo corresponde a las frecuencias de respuestas a una pregunta específica dicotómica aplicada en dos cursos de una escuela primaria (Mielke y Berry 1996). Si bien en el artículo se presentaron las respuestas clasificadas según el sexo de los estudiantes, al haberse hallado interacción entre las tablas de contingencia y el sexo, las tablas se pueden analizar por separado según el sexo de los estudiantes; por esta razón sólo se tomó en cuenta a las respuestas del género femenino para el desarrollo del ejemplo ilustrativo.

La tabla de contingencia del segundo ejemplo muestra los resultados de un estudio retrospectivo en el cual se comparan los tratamientos de cirugía y terapia radioactiva en pacientes con cáncer de laringe. La respuesta mide el estado del paciente luego de dos años de tratamiento. La misma fue provista por Agresti (2002) y tomada de Mendenhall *et al.* (1984).

El tercer ejemplo se desarrolló utilizando para ello la tabla de contingencia en la cual se reporta la frecuencia de despidos en una compañía para el primero de cuatro períodos considerados, clasificados según la edad del empleado (Kadane 1990).

Berger y Boos (1994) utilizaron la tabla de contingencia analizada en el cuarto ejemplo para ilustrar la modificación del procedimiento propuesto por ellos en el cálculo del *p*-valor exacto en la prueba no condicionada de Barnard. Los autores no reportan la naturaleza de las variables medidas en tal tabla de contingencia.

La tabla de contingencia utilizada en el quinto ejemplo fue provista por Agresti (2002) y tomada de Hansson *et al.* (1998). La misma muestra los resultados de un estudio realizado en Suecia sobre la relación entre el uso de aspirina e infarto al miocardio sobre una muestra de 1360 pacientes. La tabla muestra el número de muertes por infarto al miocardio durante los siguientes tres años de tratamiento.

A las tablas de contingencia seleccionadas se les aplicaron las pruebas de independencia de χ^2, prueba exacta de Fisher y prueba de razón de verosimilitud (Pearson 1900, Fisher 1934, Woolf 1957, Agresti y Wackerly 1977, Siegel y Castellan 1994, Good 2000, Cytel Software 2010). Para todas estas pruebas se obtuvieron los *p*-valores exactos y asintóticos utilizando el *software* estadístico StatXact 9.0, el cual calcula el *p*-valor para la metodología clásica utilizando la distribución asintótica correspondiente, y el correspondiente para la metodología exacta utilizando pruebas de permutación (Good 2000, Cytel Software 2010).

A continuación se explican con mayor detalle los métodos y la notación empleados. Sea $\mathbf{y}$ una tabla de contingencia genérica de dimensión 2×2, sea n el tamaño de la muestra, y sean las frecuencias absolutas observadas y_{ij}, $i=1,\ 2,\ \ j=1,\ 2$. Sean asimismo $f_{i.}=\sum_j y_{ij}$, el total marginal de la *i*-ésima fila, y $c_{.j}=\sum_i y_{ij}$, el total marginal de la *j*-ésima columna.

Adicionalmente se supone que la tabla de contingencia genérica $\mathbf{y}$ pertenece al conjunto de referencia $\mathbf{\Gamma}$, el cual está constituido por todas las tablas de contingencia de dimensión 2×2 tales que sus tamaños de muestra sean n y sus totales de filas y columnas sean $f_{i.}$ y $c_{.j}$, respectivamente (Good 2000, Cytel Software 2010), la Tabla 1 muestra el esquema de una tabla de contingencia genérica $\mathbf{y}$.

Tabla 1. Esquema de una tabla de contingencia genérica $\mathbf{y}$ de dimensión 2×2.

Fila/Columna	**Columna 1**	**Columna 2**	**Total Filas**
Fila 1	y_{11}	y_{12}	$f_{1.}$
Fila 2	y_{21}	y_{22}	$f_{2.}$
Total Columnas	$c_{.1}$	$c_{.2}$	n

Las pruebas estadísticas de independencia objeto de estudio de este trabajo tienen en común que basan sus decisiones en la distribución de probabilidad de una medida de discrepancia D respecto de la hipótesis nula de independencia, calculada sobre las tablas de contingencias genéricas $\mathbf{y}$, siendo $D(\mathbf{y})$ una variable aleatoria, de tal forma que se declarará significativa a la tabla de contingencia observada $\mathbf{x}$ si ocurre que $\Pr\left[D(\mathbf{y})\geq D(\mathbf{x})\right]\leq\alpha$, donde α es el nivel de significación de la prueba. El lado izquierdo de la inecuación anterior se conoce como *p*-valor, o simplemente *p*; así, una tabla de contingencia se declarará significativa siempre que $p\leq\alpha$, lo cual significa que la probabilidad de que las tablas $\mathbf{y}$ sean tan o más extremas que la tabla observada $\mathbf{x}$ es baja, favoreciendo la hipótesis alternativa de asociación entre filas y columnas de la tabla de contingencia $\mathbf{x}$, en vez de la hipótesis nula de independencia. Es pertinente aclarar que los totales marginales de filas y columnas de las tablas genéricas $\mathbf{y}$ del el conjunto de referencia $\mathbf{\Gamma}$ están definidos por los totales marginales observados en la tabla $\mathbf{x}$ (Agresti y Wackerly 1977, Good 2000, Cytel Software 2010).

Las medidas de discrepancia para las pruebas de independencia consideradas dadas por

$$\chi^2(\mathbf{y})=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(y_{ij}-f_{i.}c_{.j}/n\right)^2}{f_{i.}c_{.j}/n}$$

para la prueba de independencia de χ^2 basada en el estadístico de Pearson;

$$F(\mathbf{y})=-2\log h(\mathbf{y})-\log\left(\sqrt{2\pi f_{1.}f_{2.}c_{.1}c_{.2}/n^3}\right)$$

para la prueba exacta de Fisher, con

$$h(\mathbf{y})=\left[\binom{c_{.1}}{y_{11}}\binom{c_{.2}}{y_{12}}\right]\Big/\binom{n}{f_{1.}},$$

donde la notación $\binom{t}{u}$ representa el número de subconjuntos diferentes de *u* elementos seleccionados a partir de un conjunto de *t* elementos distintos, $t\geq u$, $t,\ u$ números naturales; y

$$V(\mathbf{y})=2\sum_{i=1}^{2}\sum_{j=1}^{2}y_{ij}\log\left(\frac{y_{ij}}{f_{i.}c_{.j}/n}\right)$$

para la prueba de razón de verosimilitud. Todas estas medidas de discrepancia convergen a la distribución asintótica χ^2 con un (1) grado de libertad en la medida que el tamaño de muestra n aumenta (Siegel y Castellan 1994, Good 2000, Cytel Software 2010).

Para comparar ambos enfoques, el clásico y el asintótico, se siguió la recomendación de Berger (2000), quien propone una estrategia basada en la medida de similitud Δ para decidir entre el uso de las pruebas exactas o asintóticas, así

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

Siendo $\Delta = \left| p_{\text{asintótico}} - p_{\text{exacto}} \right|$; $p_{\text{asintótico}}$ el *p*-valor de la prueba calculado con la metodología asintótica; p_{exacto} el *p*-valor calculado con las pruebas de permutación, y Δ^* el umbral de decisión entre los *p*-valores de ambas metodologías.

Si bien Berger (2000) no propone ningún valor específico para Δ^*, en este trabajo se tomó a $\Delta^* = 0,01$, por ser una cantidad mucho menor que el nivel de significación $\alpha = 0,05$, el cual es utilizado frecuentemente en los trabajos de investigación; de tal manera que si $\Delta \leq 0,01$ se asume que ambos enfoques convergen y por tanto se consideran equivalentes; y en caso contrario, se asume que ambos *p*-valores fueron discrepantes, prefiriéndose en ese caso el *p*-valor exacto.

RESULTADOS

Ejemplo 1

La Tabla 2A presenta tamaño de muestra $n = 31$, el cual puede considerarse pequeño, y siendo además los totales marginales de las filas y columnas balanceados.

En la Tabla 2B se observa que la prueba de Fisher presentó un *p*-valor exacto mayor que el *p*-valor asintótico, mientras que ocurrió lo contrario en las demás pruebas, siendo el valor de Δ mayor que el umbral establecido en este estudio $\Delta^* = 0,01$.

Considerando los valores obtenidos para Δ y el umbral establecido, y siguiendo la recomendación de Berger (2000), para este ejemplo en particular se deberían utilizar metodologías exactas.

Tabla 2. A. Clasificación de las respuestas a una pregunta dicotómica aplicada en dos cursos distintos en estudiantes de sexo femenino. B. *p*-valores exacto, asintótico y diferencia Δ para las pruebas de independencia aplicadas.

A. Tabla de contingencia			
Curso/Respuesta	**Sí**	**No**	**Total**
Primero	10	4	14
Cuarto	6	11	17
Total	16	15	31
B. *p*-valores y diferencia Δ			
***p*-valor**	χ^2	**Fisher**	**Verosimilitud**
p-valor exacto	0,04512570	0,07317275	0,04245759
p-valor asintótico	0,07317275	0,04860854	0,07317275
Δ	0,02804705	0,02456421	0,03071516

Fuente Tabla A: Mielke y Berry (1996).

Es interesante notar que, si se tomara una decisión al nivel de significación del 5%, las decisiones dentro de cada prueba serían contrarias según se utilizase la metodología exacta o asintótica. En concreto, se rechazaría la hipótesis nula de independencia para las pruebas de χ^2 y de razón de verosimilitud, y no se rechazaría la hipótesis de independencia para la prueba exacta de Fisher. Si tales decisiones se basaran en los *p*-valores exactos, las respuestas a la pregunta dicotómica se considerarían asociadas al curso para las pruebas de independencia de χ^2 y de razón de verosimilitud, y se considerarían independientes del curso para la prueba exacta de Fisher; asimismo, las decisiones serían las contrarias si éstas se basaran en los *p*-valores asintóticos.

Ejemplo 2

La tabla de contingencia 3A es relativamente balanceada para los totales marginales de fila y desbalanceada para los totales marginales de las columnas; presenta además tamaño de muestra pequeño $n = 41$.

En la Tabla 3B se observa que, para las tres pruebas aplicadas, el valor de las diferencias entre los *p*-valores exactos y asintóticos Δ es muy superior al umbral establecido Δ^*, a pesar de que en la tabla los totales marginales de filas son balanceados; pudiera entonces pensarse que estas diferencias entre los *p*-valores exactos y asintóticos se deben al tamaño de muestra y al desbalanceo en los marginales de las columnas, pudiendo además conjeturarse que en la medida que el tamaño de muestra n es más pequeño estas diferencias Δ pudieran ser mayores. Sin embargo, a pesar de tales diferencias, las decisiones a tomar sobre la hipótesis de independencia serían la misma para las tres pruebas consideradas, esto es, no rechazar la hipótesis nula de independencia. Es decir, el estado del paciente -cáncer controlado, cáncer no controlado- no estaría asociado con la terapia aplicada.

Tabla 3. A. Resultados de un estudio retrospectivo en el tratamiento de cáncer de laringe. B. *p*-valores exacto, asintótico y diferencia Δ para las pruebas de independencia aplicadas.

A. Tabla de contingencia

Tratamiento/ Resultado	**Cáncer controlado**	**Cáncer no Controlado**	**Total**
Cirugía	21	2	23
Terapia Radiactiva	15	3	18
Total	36	5	41

B. *p*-valores y diferencia Δ

***p*-valor**	χ^2	**Fisher**	**Verosimilitud**
p-valor exacto	0,63842578	0,63842578	0,63842578
p-valor asintótico	0,43890077	0,41569295	0,44058417
Δ	0,19952501	0,22273283	0,19784161

Fuente Tabla A: Mendenhall *et al.* (1984).

Ejemplo 3

La tabla de contingencia 4A presenta un tamaño de muestra grande $n = 249$; y es balanceada en sus totales marginales de filas, aunque fuertemente desbalanceada en los totales marginales de las columnas.

En la Tabla 4B se observa que, para las tres pruebas aplicadas, las diferencias entre los *p*-valores exactos y asintóticos Δ fueron mucho menores que el umbral Δ^* establecido y las decisiones a tomar al nivel de significación $\alpha = 0,05$ son iguales para las tres pruebas, y éstas pueden basarse en los *p-valores* asintóticos según la recomendación de Berger (2000); en este caso, tal decisión sería rechazar la hipótesis nula de independencia entre la edad del empleado y su condición de retenido o despedido por la empresa, y considerar que los despidos están asociados con la edad del trabajador.

También se observa que los *p*-valores fueron muy pequeños para todas las pruebas y por supuesto también sus diferencias Δ, lo que conlleva a conjeturar que en la medida que la hipótesis nula de independencia tienda a ser falsa, (es decir, que la retención de los empleados esté asociada con la edad de éstos), va a ser más frecuente que Δ sea inferior al Δ^* establecido.

Tabla 4. Empleados despedidos y retenidos por cierta empresa para cuatro períodos de despidos clasificados según la edad. B. *p*-valores exacto, asintótico y diferencia Δ para las pruebas de independencia aplicadas.

A. Tabla de contingencia

Edad/Decisión	**Despedidos**	**No Despedidos**	**Total**
≥ 40 años	18	129	147
< 40 años	0	102	102
Total	18	231	249

B. *p*-valores y diferencia Δ

***p*-valor**	χ^2	**Fisher**	**Verosimilitud**
p-valor exacto	0,00006748	0,00022801	0,00004941
p-valor asintótico	0,00004496	0,00024331	0,00000800
Δ	0,00002252	0,00001530	0,00004141

Fuente Tabla A: Kadane (1990).

Ejemplo 4

La tabla de contingencia 5A presenta un tamaño de muestra grande, $n = 330$, y es desbalanceada en los totales marginales de las filas y columnas.
En la Tabla 5B se observa que en la prueba exacta de Fisher y la prueba de χ^2 las diferencias Δ fueron inferiores al umbral Δ^*, mientras que en la prueba de razón de verosimilitud ocurrió lo contrario.

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

Para el caso de la prueba de razón de verosimilitud, la decisión a tomar al nivel de significación $\alpha = 0,05$ debe basarse en el *p*-valor exacto, siguiendo la recomendación de Berger (2000), por lo cual, para este ejemplo no se rechazaría la hipótesis nula de independencia entre las categorías de las filas y las columnas. Esta discrepancia en la decisión en relación a las otras dos pruebas pudiera deberse al desbalanceo presente en los totales marginales de las filas, lo cual pudiera afectar la sensibilidad de la prueba de la razón de verosimilitud (Tabla 5B).

Tabla 5. A. Tabla de contingencia utilizada por Berger y Boos (1994) para ilustrar la modificación propuesta por ellos para el cálculo del *p*-valor exacto de la prueba no condicionada de Barnard. B. *p*-valores exacto, asintótico y diferencia Δ para las pruebas de independencia aplicadas.

A. Tabla de contingencia			
Muestra/Respuesta	**Éxito**	**Fracaso**	**Total**
Muestra 1	14	33	47
Muestra 2	48	235	283
Total	62	268	330
B. *p*-valores y diferencia Δ			
***p*-valor**	χ^2	**Fisher**	**Verosimilitud**
p-valor exacto	0, 04441983	0,04441980	0,06758505
p-valor asintótico	0,03710219	0,04081388	0,04731249
Δ	0,00731764	0,00360592	0,02027256

Fuente Tabla A: Berger y Boos (1994).

Ejemplo 5

La tabla de contingencia 6A presenta un tamaño de muestra grande, $n = 1360$, y donde, si bien los totales marginales de las filas son balanceados, los totales marginales de columnas están fuertemente desbalanceados.

En la Tabla 6B se observa que para las tres pruebas aplicadas, Fisher, χ^2 y razón de verosimilitud, a pesar de que el tamaño de muestra n sea tan grande, las diferencias entre los *p*-valores exactos y asintóticos Δ fueron superiores al umbral Δ^*. Esta falta de convergencia de los *p*-valores asintóticos a los exactos pudiera deberse al fuerte desbalanceo presente en los totales marginales de las columnas.

Tabla 6. A. Resultados del estudio sueco sobre uso de aspirina e infarto al miocardio. B. *p*-valores exacto, asintótico y diferencia Δ para las pruebas de independencia aplicadas.

A. Tabla de contingencia			
Tratamiento/Resultado	**Infarto**	**No infarto**	**Total**
Placebo	28	656	684
Aspirina	18	658	676
Total	46	1314	1360
B. *p*-valores y diferencia Δ			
***p*-valor**	χ^2	**Fisher**	**Verosimilitud**
p-valor exacto	0,17678162	0,17678162	0,17678162
p-valor asintótico	0,14444339	0,14628120	0,14281592
Δ	0,03234772	0,03050042	0,03396570

Fuente Tabla A: Hansson *et al.* (1998).

DISCUSIÓN

Los resultados mostrados en la tablas 2 y 3 corroboran la afirmación de Good (2000) y Cytel Software (2010), quienes indican que en presencia de muestras pequeñas, tablas dispersas y/o desbalanceadas, los *p*-valores exactos y asintóticos muchas veces presentan un comportamiento muy disímil. Posiblemente esta discrepancia se deba a que el tamaño del conjunto de referencia $\boldsymbol{\Gamma}$ crece en la medida que aumentan tanto el tamaño de muestra como la dimensión de la tabla (Gail y Mantel 1977, Greselin 2003). Así pues, puede esperarse que para tablas de contingencia de dimensión 2×2 con tamaño entre los *p*-valores exacto y asintótico ocurría más rápido, incluso para tablas con tamaño de muestra pequeño, siendo mucho más rápido en tablas de contingencia doblemente ordenadas, posiblemente debido al tamaño mucho más grande de los conjuntos de referencia $\boldsymbol{\Gamma}$ en ambas situaciones. Asimismo, los resultados de este estudio son similares a los de Pérez *et al.* (2013) y Pérez-Ybarra *et al.* (2014), aunque la convergencia observada fue más lenta para las tablas de contingencia de dimensión 2×2.

Por otra parte, puede observarse en la Tabla 2 que los *p*-valores tanto exactos como asintóticos de las pruebas aplicadas tienden a presentar valores diferentes de una prueba a otra. Esto indica que las decisiones a tomar dependerán no sólo del enfoque seleccionado sino también de la prueba de independencia aplicada. Al respecto, Agresti (2002) señala que las distribuciones de probabilidad de las medidas de discrepancia no convergen a la distribución asintótica a la misma velocidad. Sin embargo, la no convergencia de los *p*-valores exactos y asintóticos no necesariamente implica que las decisiones serán contrarias en ambos enfoques, tal y como ocurrió en la tablas 2 y 5, sino que pueden conducir a la misma decisión, como se observó en las tablas 3, 4 y 6. Si bien el comportamiento en estas tablas fue disímil, lleva a pensar que el solo hecho de tener un tamaño de muestra pequeño y una tabla dispersa no es razón suficiente para garantizar que las decisiones a tomar sean siempre distintas. Por esta razón lo idóneo en estos casos sería utilizar el *p*-valor exacto al realizar inferencias y evitar incurrir en conclusiones discrepantes para ambos enfoques; es decir, el investigador debería ser conservador y utilizar en estos casos el *p*-valor exacto.

Cabe destacar que las tablas 2 y 5 ilustran un ejemplo de lo que pudiera ocurrir cuando los *p*-valores presentan valores cercanos al nivel de significación α, el cual para este trabajo se asumió de muestra pequeño, el conjunto de referencia $\boldsymbol{\Gamma}$ sea relativamente pequeño y por lo tanto sea más difícil que los *p*-valores exactos converjan a los asintóticos. En ese sentido, Sánchez (2005), quien condujo un estudio de simulación en tablas de dimensión 2×2 y de dimensión *R*×*C* no ordenadas, encontró que para tablas de contingencia de dimensión 2×2 la convergencia entre ambos *p*-valores ocurría más rápido para tamaños de muestra muy grandes y tablas balanceadas, lo cual coincide con lo ocurrido en estos dos ejemplos. Caso contrario fue lo encontrado por Pérez (2005), quien en un estudio de simulación en tablas de contingencia de dimensión *R*×*C* simple y doblemente ordenadas, observó que la convergencia en $\alpha = 0,05$. En dichas tablas se observa que en la cercanía del nivel de significación α, tanto en muestras grandes como pequeñas, es en donde con más frecuencia suele presentarse la discrepancia entre ambas metodologías. Bajo estas condiciones, Good (2000) recomienda inclinarse por los resultados de las pruebas de permutación, y que de ser no posible calcular el *p*-valor exacto por enumeración completa, se utilicen ensayos de Monte Carlo para su estimación aproximada (Agresti *et al.* 1979, Good 2000).

El comportamiento observado en la Tabla 4 puede deberse a dos razones:

a) El tamaño de conjunto de referencia $\boldsymbol{\Gamma}$ crece geométricamente según aumenta el tamaño de la muestra y la dimensión de la tabla de contingencia (Gail y Mantel 1977, Greselin 2003, Cytel Software 2010). Cuando el tamaño del conjunto de referencia es muy grande, la distribución de la medida de discrepancia $D(\mathbf{y})$, obtenida por permutaciones de la tabla de contingencia observada $\mathbf{x}$, presenta muchos puntos y tiende a ser muy parecida a la distribución asintótica correspondiente, de allí que los *p*-valores exactos y asintóticos tiendan a ser muy parecidos entre sí.

b) Como hay razones para pensar que la hipótesis nula de independencia entre las categorías de filas y columnas es falsa en el ejemplo 3, ya que los *p*-valores de observados son muy pequeños, pudiera estar ocurriendo lo siguiente: cuando la hipótesis nula de independencia es falsa, las tablas de contingencia consideradas para el cálculo del *p*-valor tenderán a presentar medidas de discrepancia $D(\mathbf{y})$ cercanas o mayores que el valor crítico:

$$d_\alpha(\mathbf{y}) = \min\left\{d(\mathbf{y}) : \Pr\left[D(\mathbf{y}) \geq d(\mathbf{y})\right] \leq \alpha\right\}$$

y por ende, los *p*-valores correspondientes a las tablas de contingencia observadas tenderán también a presentar valores cercanos o menores que el nivel de significación α ; 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. Bajo tal escenario, puede esperarse que las diferencias Δ entre ambos *p*-valores será bastante pequeña y que no se superará el umbral Δ^* para tamaños de muestra relativamente moderados o pequeños, y más aún para tamaños de muestra grande, presentándose convergencia de los *p*-valores de asintóticos a los exactos. Este mismo comportamiento fue observado por Pérez *et al.* (2013) y Pérez-Ybarra *et al.* (2014) en tablas de contingencia de dimensión $R\times C$ tanto no ordenadas como simplemente ordenadas.

En el caso particular de la Tabla 6, en la cual dado el tamaño de muestra, pudiera esperarse que los *p*-valores de ambos enfoques fueran convergentes con valor de $\Delta^* < 0,01$, sin embargo, esto no ocurrió. Este comportamiento pudiera deberse al hecho de la presencia de desbalanceo en filas y columnas, el cual, tal como afirma Good (2000), tiende a hacer más lenta la tasa de convergencia de los *p*-valores asintóticos a los exactos, aún en presencia de un tamaño de muestra grande, por lo cual la recomendación generalizada de utilizar la distribución asintótica si el valor esperado de las celdas es mayor a 5 (Siegel y Castellan 1994), no es razón suficiente para dejar de lado a la metodología exacta, siempre que sea posible calcular el *p*-valor correspondiente.

CONCLUSIONES

Los ejemplos analizados mostraron que el comportamiento de las metodologías exacta y asintótica y por ende, su elección, depende del tamaño de muestra, dimensión, balanceo y dispersión de la tabla de contingencia analizada, lo cual coincide con otros trabajos publicados y ratifica el comportamiento diferencial de los enfoques asintótico y exacto.

Aunque la metodología asintótica puede aplicarse en casos de tablas de contingencia balanceadas y con tamaños de muestra grande, se recomienda basar las conclusiones de los trabajos de investigación en los *p*-valores exactos, siempre que sea posible calcularlos, ya que no hay garantía de que los *p*-valores de ambos enfoques sean siempre convergentes, y por tanto, equivalentes.

Las conclusiones de este trabajo son válidas para cualquier tabla de contingencia de dimensión 2×2 aplicada en cualquier disciplina o área del conocimiento, por cuanto que las mismas son inherentes a la naturaleza de las pruebas de independencia aplicadas sobre las tablas y a las metodologías exacta y asintótica, y no a la disciplina dentro de la cual fueron construidas u obtenidas.

REFERENCIAS BIBLIOGRÁFICAS

AGRESTI A. 2002. Categorical data analysis. Wiley-Interscience, New Jersey, USA, pp. 717.

AGRESTI A. 2007. An introduction to categorical data analysis. Wiley-Interscience, New Jersey, USA, pp. 372.

AGRESTI A, WACKERLY D. 1977. Some exact conditional tests of independence for $R\times C$ cross-classification tables. Psychometrika. 42(1):111-125.

AGRESTI A, WACKERLY D, BOYETT JM. 1979. Exact conditional tests for cross-classifications: approximation of attained significance levels. Psychometrika. 44(1):75-83.

BERGER VW. 2000. Pros and cons of permutation tests in clinical trials. Stat Med. 19(10):1319-1328.

BERGER RL, BOOS DD. 1994. *P* values maximized over a confidence set for the nuisance parameter. J. Am. Stat. Assoc. 89(427):1012-1016.

CYTEL SOFTWARE. 2010. Statxact 9 with Cytel studio. Statistical software for exact nonparametric inference. User manual. Cytel Software, New York, USA, pp. 1345.

DANIEL WW. 2008. Bioestadística. Base para el análisis de las ciencias de la salud. Limusa-Wiley, México, pp. 915.

FISHER RA. 1934. Statistical methods for research workers. Oliver and Boyd, Edinburgh, UK, pp. 324.

FORTIN MJ, JACQUEZ GM, SHIPLEY B. 2002. Computer-intensive methods. *In*: EL-SHAARAWI AH, PIEGORSCH WW. Encyclopedia of Environmetrics. John Wiley and Sons, Chichester, pp. 399-400.

GAIL M, MANTEL N. 1977. Counting the number of *R*×*C* contingency tables with fixed margins. J. Am. Stat. Assoc. 72(360):859-862.

GOOD P. 2000. Permutation tests. A practical guide to resampling methods for testing hypotheses. Second edition. Springer-Verlag, New York, 270 p.

GRESELIN F. 2003. Counting and enumerating frequency tables with given margins. Statistica & Applicazioni. 1(2):87-104.

HANSSON L, ZANCHETTI A, CARRUTHERS SG, DAHLÖF B, ELMFELDT D, JULIUS S, MÉNARD J, RAHN KH, WEDEL H, WESTERLING S. 1998. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. Lancet. 351(9118):1755-1762.

KADANE JB. 1990. A statistical analysis of adverse impact of employer decisions. J. Am. Stat. Assoc. 85(412):925-933.

MENDENHALL WM, MILLION RR, SHARKEY DE, CASSINI NJ. 1984. Stage T_3 squamous cell carcinoma of the glottic larynx treated with surgery and/or radiation therapy. Int. J. Radiat. Oncol. Biol. Phys. 10:357-363.

MIELKE PW, BERRY KJ. 1996. Exact probabilities for first-order and second-order interactions in 2×2×2 contingency tables. Educ. Psychol. Meas. 56(5):843-847.

PEARSON K. 1900. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos Mag Ser 5. 50(302):157-175.

PÉREZ LM. 2005. Comparación de las pruebas de permutación y clásicas (asintóticas) aplicadas en tablas de contingencia de dimensión R×C simple y doblemente ordenadas. Maracay: Universidad Central de Venezuela, Facultad de Agronomía, Postgrado en Estadística [Disertación Maestría en Estadística], pp.125.

PÉREZ LM, SÁNCHEZ L, MARTÍNEZ H, QUINTANA W. 2013. Comparación de las pruebas de permutación y asintóticas aplicadas en tablas de contingencia de dimensión r×c no ordenadas utilizadas en estudios biomédicos. Salus. 17(1):61-70.

PÉREZ-YBARRA LM, SÁNCHEZ-AGUILERA LA, MARTÍNEZ H, QUINTANA-RIVERO WF. 2014. Comparación de las pruebas de permutación y asintóticas aplicadas en tablas de contingencia de dimensión R×C simplemente ordenadas utilizadas en estudios biomédicos. Acta Cien. Ven. 65(1):40-51.

SÁNCHEZ LA. 2005. Comparación de los métodos permutacional y asintótico en pruebas de hipótesis aplicadas a tablas de contingencia de dimensión 2×2 y R×C no ordenadas. Maracay: Universidad Central de Venezuela, Facultad de Agronomía, Postgrado en Estadística [Disertación Maestría en Estadística], pp. 135.

SIEGEL S, CASTELLAN NJ. 1994. Estadística no paramétrica aplicada a las ciencias de la conducta. Editorial Trillas, México, pp. 437.

WOOLF B. 1957. The log likelihood ratio test (The G-test). Ann. Hum. Genet. 21(4):397-409.