

Contraste de independencia con corrección de continuidad asimétrica: Una aplicación al turismo cultural.

Montero Lorenzo, Jose Maria (Jose.MLonrenzo@uclm.es)
Fernández-Avilés Calderón, Gema (gema.faviles@uclm.es)
Departamento de Estadística
Universidad de Castilla La Mancha

García Centeno, María del Carmen (garcen@ceu.es)
Departamento de Métodos Cuantitativos e Informática
Universidad CEU San Pablo

RESUMEN

La utilización de correcciones de continuidad en el estadístico Chi-cuadrado, a la hora de contrastar la hipótesis de independencia entre los niveles de dos o más factores, es una de las cuestiones clásicas de la estadística cualitativa. En este artículo se propone una corrección de continuidad asimétrica para tablas de contingencia 2x2, que mejora sensiblemente la operatividad de otras correcciones clásicas como las de Yates, Cochran o Mantel. La decisión sobre si utilizar o no correcciones de continuidad, o en este último caso, cuál de ellas utilizar, no es inocua. Ello se ilustra en el caso del turismo cultural en la ciudad de Toledo a la hora de determinar si en la conocida Ciudad Patrimonio de la Humanidad se verifica o no la que hemos dado en denominar versión positiva de la Teoría de la Asimilación; esto es, si en el marco del paradigma “expectativas-confirmación de las mismas” las expectativas constituyen un elemento conductor matizado de la valoración de la satisfacción.

Palabras claves: Corrección de continuidad; tabla de contingencia; teoría de la asimilación; satisfacción; turismo.

Área temática: Métodos Cuantitativos y Turismo.

XVIII Jornadas ASEPUMA – VI Encuentro Internacional

Anales de ASEPUMA nº 18: **801**

ABSTRACT

The use of continuity correction in the chi-square statistic for testing the independence hypothesis between two or more factors, is one of the classical question in qualitative statistics. In this paper we suggest an asymmetric continuity correction for 2x2 contingency tables that improves other classical corrections such as the proposed by Yates, Cochran or Mantel. The decision about to use or not continuity corrections, or in the last case which of them to use, it is, not innocuous. This is showed in a cultural tourism case, the case of Toledo (Spain), when it comes to check whether what we call the positive version of the assimilation theory is verified in this UNESCO World Heritage city; that is to say, to check whether the expectations are a qualified driver of satisfaction in the framework of the expectation/disconfirmation paradigm.

Keywords: Continuity correction; 2x2 contingency tables; assimilation theory; satisfaction; tourism.

1. INTRODUCCIÓN

Es por todos conocido que la contrastación de la hipótesis de independencia entre dos o más factores, con dos o más niveles cada uno, es una cuestión estadística clave en muchas disciplinas científicas, pero sobre todo en aquellas en las que predomina la información de carácter cualitativo o susceptible de ordenación.

El instrumento convencional para determinar estadísticamente si dos factores son independientes o, por el contrario, están asociados, no es otro que el conocido contraste chi-cuadrado de independencia. Ahora bien, aunque se hace un uso indiscriminado de tal contraste, se ha de poner de manifiesto que la distribución del estadístico de contraste es una distribución aproximada que resulta de utilidad para grandes conjuntos de datos, pero no cuando la información de la que se dispone, como ocurre en numerosas ocasiones en la práctica real, es ciertamente escasa.

Y es que, para empezar, en el contraste de independencia entre dos factores o atributos hay que tener muy presente el procedimiento de muestreo, ya que éste constituye una pieza clave a la hora de determinar si la muestra de la que dispone el investigador pertenece a la región crítica o de rechazo de la hipótesis de independencia o no. Evidentemente, a efectos probabilísticos e inferenciales no es lo mismo que los totales marginales de los dos factores de la tabla bifactorial sean fijos, que únicamente sean fijos los de uno de ellos, que ambos sean libres de variar siendo el tamaño total de la muestra la única cantidad fijada de antemano, que ni siquiera sea fijo dicho total muestral, o que estén fijadas las frecuencias de las celdas correspondientes a un determinado nivel de uno de los factores.

Al ser el contraste de independencia un contraste de significación, para contrastar la hipótesis nula H_0 : "Los dos factores de la tabla de contingencia bidimensional son independientes" frente a la hipótesis alternativa H_1 : "Existe asociación entre ambos", se debe calcular la probabilidad (bajo la hipótesis nula) de obtener la disposición de frecuencias observada (tabla observada) o aquéllas otras que evidencien igual o mayor alejamiento de la hipótesis de independencia que la disposición de frecuencias observada, de entre todas las tablas ($R \times C$) que obedezcan al mismo diseño que la tabla observada. Una vez calculada dicha probabilidad (denominada "probabilidad exacta", Ruiz-Maya et al.,

1990 y 1995) se comparará con el nivel de significación prefijado, para determinar si la tabla observada proporciona evidencia suficiente en contra de la hipótesis de independencia formulada. Como puede observarse, el procedimiento expuesto es harto laborioso por cuanto exige la localización de las tablas que se alejan más que la observada de la hipótesis de independencia y el cálculo de la probabilidad de ocurrencia de cada una de ellas. Es por ello que no resulta extraño, dado el problema computacional expuesto, el recurso a procedimientos aproximativos como el basado en el estadístico chi-cuadrado.

El problema que surge es que la probabilidad exacta, cuyo cálculo se lleva a cabo con distribuciones discretas de probabilidad (hipergeométrica, binomial bivalente, multinomial, Poisson o binomial negativa bivalente, dependiendo del procedimiento de muestreo que se haya llevado a cabo) se aproxima a través de una distribución continua de probabilidad (la distribución Chi-cuadrado). Tradicionalmente se ha considerado, en el caso de tablas de contingencia de orden (2x2), que dicha aproximación no es suficientemente buena si alguna de las estimaciones de las frecuencias esperadas, calculadas bajo la hipótesis de independencia, es inferior a 5 y se ha señalado la necesidad de aplicar algún elemento corrector. Es en este punto donde surgen las denominadas correcciones de continuidad para intentar compensar los desajustes que tienen lugar cuando la distribución de probabilidad de las frecuencias observadas, que es discreta, es aproximada por otra de carácter continuo.

Dicho lo anterior, surge la siguiente cuestión: ¿La manera de corregir de continuidad es independiente del diseño del experimento? o, por el contrario, ¿dependiendo de cual sea el modelo la corrección de continuidad se lleva a cabo de una u otra manera? La respuesta es que diferentes diseños exigen diferentes correcciones de continuidad.

Realizadas las anteriores consideraciones de carácter teórico, las investigaciones relativas a la satisfacción del turista son un caso particularmente relevante de uso necesario del contraste de independencia. Cuestiones como ¿la satisfacción del turista está asociada con tal o cual factor? ¿Está asociada con sus expectativas previas? ¿Está asociada con la importancia que el turista asigna a los distintos servicios turísticos?, entre otras, exigen de la estadística cualitativa, más que de la estadística cuantitativa (que, por otra parte, es la base de los modelos Expectancy/Disconfirmation).

Ahora bien, en el ámbito turístico, como en muchos otros ámbitos de la realidad, el número de cuestionarios de los que dispone el investigador no suele ser tan amplio como el deseado, y en caso de que así fuese, la no respuesta a determinadas cuestiones del mismo, o la incorrecta manera de contestar (que invalida la respuesta), llevan en muchas ocasiones a que el volumen de información disponible para determinar estadísticamente si dos factores están o no asociados sea ciertamente escaso.

En estas tesituras, más habituales de lo deseado, es en las que se propone un contraste que incluye una corrección de continuidad asimétrica que supera notablemente los hasta ahora expuestos en la literatura estadística sobre el tema. Se trata de un aspecto importante cuando se trabaja con cuestionarios y, sobre todo, de uno de los innumerables tópicos aun permanecen abiertos en el campo de la estadística cualitativa. Si bien fue estudiado en profundidad a mediados del siglo pasado, las soluciones que se aportaron no pueden considerarse satisfactorias. Tal es así que Conover, 1968, entre otros autores, incluso llegó a recomendar su no utilización, argumentando que en vez de corregir el error cometido en la continuización, lo aumentan.

Tras esta sección introductoria, la sección segunda discurre por las principales correcciones de continuidad que se incluyen en el estadístico de contraste de la hipótesis de independencia para aproximar el p-valor. En la tercera se articula una propuesta de corrección de continuidad asimétrica que supera a las expuestas en la sección precedente. En la cuarta se pone en funcionamiento dicha corrección en el marco del turismo cultural. El artículo se da por finalizado con la exposición de las conclusiones más relevantes que se derivan del mismo.

2. CONTRASTE DE HIPÓTESIS DE INDEPENDENCIA DE CONTINGENCIA (2X2).

Dado que la situación más habitual a la que se enfrenta el investigador en la práctica es aquélla en que se consideran dos factores, cada uno de ellos con dos niveles, nos centraremos en el marco de las tablas de contingencia (2x2) con totales marginales fijos (también el diseño de muestreo más popular).

En la base del contraste Chi-cuadrado de independencia figura la probabilidad aproximada de obtener, bajo la hipótesis nula de independencia, la tabla observada o aquéllas otras con igual o mayor alejamiento (en cualquier dirección) de la hipótesis nula, es decir:

$$P\left(\left|N_{ij} - \hat{E}_{ij}\right| \geq n_{ij} - \hat{E}_{ij}\right) \quad \forall i, j,$$

donde n_{ij} y \hat{E}_{ij} son, respectivamente, las frecuencias observada y esperada en la celda $\{ij\}$ y N_{ij} , la frecuencia con carácter aleatorio, sigue una distribución de probabilidad hipergeométrica —la probabilidad de la tabla T_i viene dada por $P_i = \frac{n_{11}!n_{21}!n_{12}!n_{22}!}{n!n_{11}!n_{12}!n_{21}!n_{22}!}$ —

por estar fijados los totales marginales de los dos factores involucrados en la tabla bifactorial. Dicha probabilidad aproximada, que se comparará con el nivel de significación prefijado se calcula a través de:

$$P\left(\chi_1^2 \geq \frac{(n-1)(n_{11}n_{22} - n_{12}n_{21})^2}{n_{11}n_{21}n_{12}n_{22}}\right)$$

Si bien el procedimiento anterior simplifica sobremanera la realización del contraste, no es menos cierto que se está cometiendo un "error por continuización". Para corregirlo, Yates, 1934, propuso una corrección de continuidad, simétrica, que se opera en el estadístico χ_{adj} .

$$\chi_{adj}^2 = \frac{(n-1)(n_{11}n_{22} - n_{12}n_{21})^2}{n_{11}n_{21}n_{12}n_{22}} = \frac{n-1}{n} \sum_i \sum_j \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

y que consiste en restar 0,5 al valor absoluto de las diferencias $(n_{ij} - \hat{E}_{ij})$, obteniéndose una aproximación a la "probabilidad exacta", ya corregida de continuidad, mediante cualquiera de las tres expresiones siguientes:

$$P \left[\chi_1^2 \geq \frac{(n-1) \sum_i \sum_j \left(|n_{ij} - \hat{E}_{ij}| - 0,5 \right)^2}{n_1 n_2 n_{11} n_{12}} \right]$$

$$P \left[\chi_1^2 \geq \frac{(n-1) \left(|n_{11} n_{22} - n_{12} n_{11}| - 0,5n \right)^2}{n_1 n_2 n_{11} n_{12}} \right]$$

$$P \left[\chi_1^2 \geq \frac{n^2 (n-1)}{n_1 n_2 n_{11} n_{12}} \left(|n_{ij} - \hat{E}_{ij}| - 0,5 \right)^2 \right]$$

Existe abundante bibliografía relativa a la conveniencia de la utilización de la corrección de continuidad de Yates, destacando las aportaciones clásicas de Mantel y Greenhouse (1968), Mantel (1974), Fleiss (1973), Conover (1968, 1974) y Haber (1980), y las más recientes de Marín et al. (2005) y Haviland (2007).

Cochran, 1942, ante las deficiencias de la corrección de Yates, propuso tomar la frecuencia observada a la que corresponde la estimación de la frecuencia esperada más pequeña bajo la hipótesis de independencia, y distinguir dos casos: a) $n_{ij} \leq 2\hat{E}_{ij}$ y b) $n_{ij} > 2\hat{E}_{ij}$

a) $n_{ij} \leq 2\hat{E}_{ij}$:

Denomínese $\chi_{adj,0}^2$ al estadístico muestral, y $\chi_{adj,1}^2$ al máximo valor del estadístico χ_{adj}^2 que no supere el valor tomado en la muestra, $\chi_{adj,0}^2$. Se define la media aritmética de las raíces cuadradas de ambos

$$\chi_{adj,C1}^2 = \frac{\sqrt{\chi_{adj,0}^2} + \sqrt{\chi_{adj,1}^2}}{2}$$

y se aproxima la probabilidad exacta mediante $P(\chi_1^2 \geq \chi_{adj,C1}^2)$.

b) $n_{ij} > 2\hat{E}_{ij}$:

En este caso se define $\chi_{adj,1}^2$ como

$$\chi_{adj,1}^2 = \frac{n^2 (n-1)}{n_1 n_2 n_{11} n_{12}} \left(|n_{ij} - \hat{E}_{ij}| - 1 \right)^2$$

y se aproxima la probabilidad exacta a través de $\frac{1}{2}P(\chi_1^2 \geq \chi_{adj,C1}^2)$.

Una segunda aportación de Cochran propone definir el estadístico Chi-cuadrado ajustado, corregido de continuidad por la segunda corrección de Cochran, como

$$\chi_{adj,C2}^2 = \frac{\chi_{adj,0}^2 + \chi_{adj,1}^2}{2}$$

aproximándose la probabilidad exacta por $P(\chi_1^2 \geq \chi_{adj,C2}^2)$ en el caso a) y por $\frac{1}{2}P(\chi_1^2 \geq \chi_{adj,C2}^2)$ en la tesitura b).

Por último queremos destacar la corrección de continuidad de Mantel, 1976. Se distinguen nuevamente dos casos: a) $n_{ij} \leq 2\hat{E}_{ij}$ y b) $n_{ij} > 2\hat{E}_{ij}$, siendo n_{ij} la frecuencia observada a la que corresponde la estimación de la frecuencia esperada más pequeña bajo la hipótesis de independencia, \hat{E}_{ij} .

a) $n_{ij} \leq 2\hat{E}_{ij}$:

Se procede primeramente a la aproximación de la probabilidad exacta mediante el estadístico Chi-cuadrado ajustado corregido con la corrección de continuidad de Yates. A dicha aproximación le denominará P_Y :

$$P_Y = P(\chi_1^2 \geq \chi_{adj,Y}^2)$$

A continuación se define $DF_{ij} = |n_{ij} - \min \hat{E}_{ij}|$ y se denomina DF_{ij}^* a la diferencia que más se acerca, por exceso, a la que se ha tomado de referencia. Por último, se calcula

$$P_Y^* = P\left(\chi_1^2 \geq \frac{n^2(n-1)}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}\left(DF_{ij}^* - \frac{1}{2}\right)^2\right)$$

y se aproxima la probabilidad exacta por la media aritmética de P_Y y P_Y^* :

$$P_M = \frac{P_Y + P_Y^*}{2}$$

b) $n_{ij} > 2\hat{E}_{ij}$:

En este caso la manera de proceder es mucho más sencilla, puesto que se aproxima la probabilidad exacta por la mitad de la probabilidad P_Y : $P_M = \frac{P_Y}{2}$.

3. UNA CORRECCIÓN DE CONTINUIDAD ASIMÉTRICA PARA EL CONTRASTE DE INDEPENDENCIA EN TABLAS DE CONTINGENCIA (2X2).

Nuestra propuesta está basada en Montero (2002) y constituye un enfoque asimétrico de las correcciones de continuidad. Como puede verse en la referencia citada, dicha corrección corrige mejor las infraestimaciones y sobrestimaciones de la probabilidad exacta a las que conduce la corrección de Yates que lo que lo hacen las anteriormente expuestas correcciones de Cochran y Mantel. La exposición de la corrección de continuidad asimétrica que se propone se lleva a cabo en el marco de las tablas de contingencia (2x2), pero es fácilmente generalizable a tablas (RxC).

Sean n_{ij} y \hat{E}_{ij} las frecuencias observada y esperada bajo la hipótesis de independencia, respectivamente, en la celda $\{ij\}$ de una tabla de contingencia (2x2). Entonces,

$$P_{exact} \approx \begin{cases} P\left(N_{ij} \leq n_{ij} + \frac{1}{2}\right) + P\left(N_{ij} \geq 2\hat{E}_{ij} - n_{ij} + \Delta\right) & \text{si } n_{ij} \leq \hat{E}_{ij} \\ P\left(N_{ij} \geq 2\hat{E}_{ij} - n_{ij} + \Delta\right) + P\left(N_{ij} \leq n_{ij} + \frac{1}{2}\right) & \text{si } n_{ij} > \hat{E}_{ij} \end{cases}$$

donde

$$N_{ij} \xrightarrow{\text{aprox}} N\left(\frac{n_{i\cdot}n_{\cdot j}}{n}, \sqrt{\frac{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}{n^2(n-1)}}\right)$$

Es decir,

$$P_{exact} \approx \begin{cases} P\left(\xi^* \leq \frac{\hat{D}_{ij} + \frac{1}{2}}{\sqrt{\frac{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}{n^2(n-1)}}}\right) + P\left(\xi^* \geq \frac{\Delta - \hat{D}_{ij}}{\sqrt{\frac{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}{n^2(n-1)}}}\right) & \text{si } n_{ij} \leq \hat{E}_{ij} \\ P\left(\xi^* \leq \frac{\Delta - \hat{D}_{ij}}{\sqrt{\frac{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}{n^2(n-1)}}}\right) + P\left(\xi^* \geq \frac{\hat{D}_{ij} - \frac{1}{2}}{\sqrt{\frac{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}{n^2(n-1)}}}\right) & \text{si } n_{ij} > \hat{E}_{ij} \end{cases}$$

donde ξ^* es a distribución normal estándar, $\hat{D}_{ij} = n_{ij} - \hat{E}_{ij}$, y Δ es una cantidad que se calcula como sigue:

1. Tómesese la parte entera de $2\hat{E}_{ij} - n_{ij}$.
2. Súmese 0,5 a la cantidad anterior y denótese el resultado por θ .
3. Obténgase Δ de la ecuación $\Delta = \theta - 2\hat{E}_{ij} + n_{ij}$.

Es necesario considerar las siguientes dos excepciones a la regla anterior:

(i) En caso de que $2\hat{E}_{ij} - n_{ij} \leq \min(n_{11}, n_{21}, n_{12}, n_{22})$ se tiene que $P\left(\xi^* \geq \frac{\Delta - \hat{D}_{ij}}{\sqrt{\frac{n_{11}n_{21}n_{12}n_{22}}{n^2(n-1)}}}\right) = 0$.

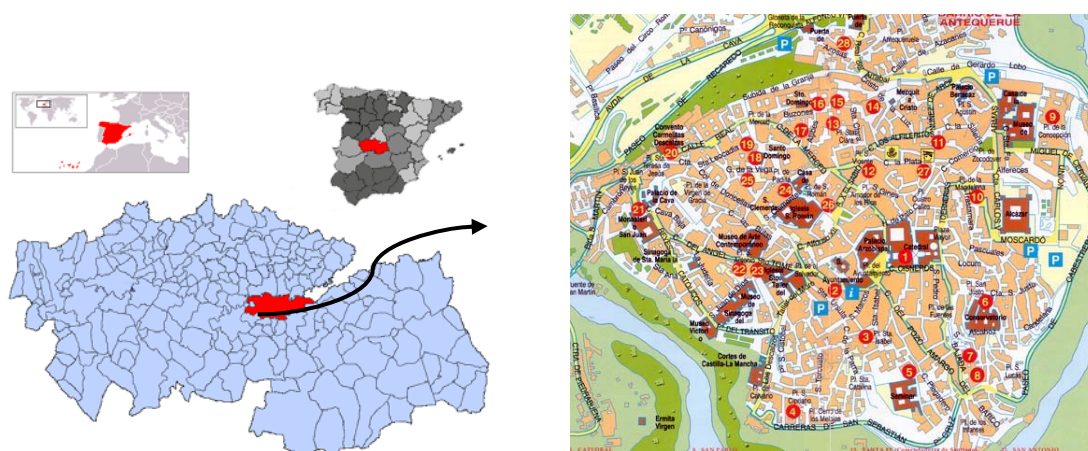
(ii) Cuando $2\hat{E}_{ij} - n_{ij} \leq 0$, $P\left(\xi^* \leq \frac{\Delta - \hat{D}_{ij}}{\sqrt{\frac{n_{11}n_{21}n_{12}n_{22}}{n^2(n-1)}}}\right) = 0$.

Finalmente, el test de independencia se lleva a cabo comparando la probabilidad correspondiente (depende del caso) con el nivel de significación prefijado.

4. UNA APLICACIÓN EN EL ÁMBITO DEL TURISMO CULTURAL.

Para ilustrar el procedimiento expuesto en la sección anterior se ha distribuido en el casco histórico de Toledo un cuestionario relativo a las expectativas y satisfacción del turista con un elenco de 20 servicios turísticos. Como es bien conocido, Toledo es Ciudad Patrimonio de la Humanidad y tiene una muy elevada dependencia del sector turístico. En concreto, de acuerdo con Esteban et al. (2005), el sector servicios acapara el 84,5% del valor añadido bruto que se genera en el casco histórico de la ciudad, la mitad del mismo correspondiente a actividades turísticas. El 45% de los visitantes de la ciudad son extranjeros, y en términos de gasto en la ciudad, alrededor del 80% manejan un presupuesto por persona y día entre 50 y 100 euros, mientras que un 14% gasta entre 100 y 200 euros. La ciudad alberga 11.000 establecimientos comerciales.

Figura 1. Toledo: Una Ciudad Patrimonio de la Humanidad



La información relativa a la satisfacción del turista con los servicios ofrecidos por la ciudad se recogió durante los meses de abril, mayo y junio de 2009. Se reunieron un total 1.500 cuestionarios si bien, dadas la dificultad de los turistas a la hora de hacer patente si sus expectativas eran superiores o no a la satisfacción obtenida, el número de cuestionarios válidos para contrastar la por nosotros denominada versión positiva de la Teoría de la Asimilación se redujo notablemente.

El cuestionario incluía los siguientes atributos de carácter turístico: alojamiento, atractivo turístico y/o paseos, entretenimiento y actividades culturales recreativas, gastronomía, comercio, limpieza, higiene y baños públicos, comunicaciones/telefonía/Internet, alumbrado publico, hospitalidad/trato recibido, información, seguridad, oficinas de información turística, prestadores de servicios turísticos, transporte público de media y larga distancia, carteles y señalización, bancos/cajeros automáticos, transporte local, cuidado ambiental, accesibilidad, y capacidad para resolver problemas.

El objetivo que nos planteamos, como ya de adelantó, es determinar si en el turismo cultural de la Ciudad de Toledo puede suponerse el cumplimiento de la Teoría de la Asimilación en la versión que hemos dado en denominar “positiva”. La Teoría de la Asimilación (Sherif y Hovland, 1961, y Oliver, 1997) sugiere que los individuos sufren un conflicto psicológico cuando perciben discrepancias entre la satisfacción obtenida por un producto o servicio y sus expectativas previas sobre el mismo, de tal manera que tienden a ajustar ambas, e incluso a eliminar la diferencia entre ellas. En estas circunstancias, las expectativas constituyen un elemento conductor de la valoración de la satisfacción. La versión positiva de la Teoría de la Asimilación hace referencia a la resolución del conflicto psicológico de los individuos por la vía de una moderada exacerbación, de tal manera que cuando las expectativas son elevadas las corrigen someramente al alza y cuando son bajas la corrección es, también moderada, pero a la baja.

En términos estadísticos el contraste del cumplimiento de la versión positiva de la Teoría de la Asimilación se concreta en contrastar la hipótesis de independencia entre los niveles de los factores “Expectativas” (bajas, altas) y “Disconfirmación” (negativa, positiva). A estos efectos, se entiende por disconfirmación positiva una escasa superación de las expectativas y por disconfirmación negativa el no alcance, por poco, de las mismas.

El paso clave es la contrastación de dicha hipótesis, si bien el paso definitivo es, en caso de rechazo de la misma, determinar si existe asociación positiva entre ambos factores (expectativas altas-disconfirmación positiva y expectativas bajas-disconfirmación negativa). En caso de que exista no se rechaza la Teoría, lo cual implica una leve exacerbación de las expectativas. En caso de que la asociación sea negativa (corrección de las expectativas) se podría afirmar que en caso de elevadas expectativas las cosas no son tan buenas como se esperaba, pero en caso de pocas expectativas las cosas no son tan malas como era de prever.

En la Tabla 1 se puede ver el resultado de la satisfacción del turista para los veinte servicios considerados en el cuestionario.

Tabla 1. Satisfacción del turista

Servicio	Ponderación	Puntuación
	λ_i	\bar{P}
Alojamiento	0,061	6,950
Atractivo turístico y/o paseos	0,102	7,683
Entretenimiento y actividades culturales recreativas	0,073	6,907
Gastronomía	0,073	7,390
Comercio	0,047	6,477
Limpieza, higiene, baños públicos	0,051	6,301
Comunicaciones/ Telefonía/ Internet/	0,040	6,127
Alumbrado público	0,036	6,483
Hospitalidad/Trato recibido	0,054	6,937
Calidad de la información	0,048	6,863
Seguridad	0,050	6,809
Oficinas de información turística	0,047	6,560
Prestadores de servicios turísticos	0,045	6,363
Transporte de media y larga distancia	0,041	6,577
Carteles y señalización	0,037	6,303
Bancos/Cajeros automáticos	0,037	6,303
Transporte local	0,039	6,250
Cuidado ambiental	0,040	6,567
Accesibilidad	0,041	6,277
Capacidad para resolver problemas	0,036	6,120
Satisfacción Global		7,150

Fuente: elaboración propia.

$0 \leq \lambda_i \leq 1$, representa la importancia que para el turista tiene el servicio i -ésimo.

En la Tabla 2 se pueden ver los principales resultados relativos al contraste de la hipótesis de independencia entre la satisfacción con los atributos considerados y el tipo de disconfirmación (leve) utilizando la versión del estadístico Chi-cuadrado que incorpora una corrección de continuidad asimétrica, así como los respectivos valores de la Q de Yule. Como puede apreciarse, en todos los casos, salvo en entretenimiento y actividades culturales recreativas, alumbrado público y carteles y señalización, puede suponerse la actuación de la versión positiva de la Teoría de la Asimilación (nivel de significación del 5%); es decir, el p -valor obtenido con el estadístico propuesto es inferior al 5% y la Q de Yule es positiva. Se ha incluido un factor de confirmación de asociación positiva que se ha calculado como la ratio “porcentaje de turistas con reducidas o elevadas expectativas/porcentaje de turistas que respondieron el cuestionario”, para evitar casos en los que la asociación positiva esté basada en un pequeño porcentaje de cuestionarios cumplimentados con elevadas expectativas.

Tabla 2. Contraste de independencia y medición de la asociación (con un estadístico Chi-cuadrado que incluye una corrección de continuidad asimétrica)

Atributo	$n_{11} > \hat{E}_{11}$ (todos los casos)			Q de Yule	Confirmación de la Teoría de la Asimilación (versión positiva)	Factor de Confirmación
	P(1)	P(2)	Probabilidad exacta			
Alojamiento	0,0000	0,0000	0,0000*	0,91	Sí	0,76
Atractivo turístico y/o paseos	0,0000	0,0000	0,0000*	0,61	Sí	0,81
Entretenimiento y actividades culturales recreativas	0,0901	0,0025	0,1511	0,11	No	0,69
Gastronomía	0,0000	0,0000	0,0000*	0,52	Sí	0,86
Comercio	0,0006	0,0010	0,0016*	0,34	Sí	0,74
Limpieza, higiene, baños públicos	0,0150	0,0174	0,0324*	0,24	Sí	0,63
Comunicaciones/ Telefonía/ Internet/ Alumbrado público	0,0011	0,0019	0,0030*	0,35	Sí	0,57
Hospitalidad/Trato recibido	0,0793	0,0869	0,1662	0,16	No	0,74
Calidad de la información	0,0000	0,0000	0,0000*	0,59	Sí	0,79
Seguridad	0,0026	0,0025	0,0051*	0,31	Sí	0,79
Oficinas de información turística	0,0000	0,0000	0,0000*	0,48	Sí	0,71
Prestadores de servicios turísticos	0,0019	0,0020	0,0039*	0,30	Sí	0,85
Transporte de media y larga distancia	0,0000	0,0000	0,0000*	0,87	Sí	0,32
Carteles y señalización	0,0009	0,0013	0,0022*	0,29	Sí	0,67
Bancos/Cajeros automáticos	0,0256	0,0287	0,0543	0,21	No	0,65
Transporte local	0,0122	0,0192	0,0314*	0,24	Sí	0,69
Cuidado ambiental	0,0000	0,0000	0,0000*	0,46	Sí	0,68
Accesibilidad	0,0011	0,0011	0,0022*	0,31	Sí	0,72
Capacidad para resolver problemas	0,0006	0,0011	0,0017*	0,36	Sí	0,73
	0,0000	0,0004	0,0004*	0,43	Sí	0,61

Fuente: elaboración propia.

* Significativo al 5%.

$$p^{(1)} = P\left(N_{ij} \leq 2\hat{E}_{11} - n_{11} + \Delta\right); \quad p^{(2)} = P\left(N_{11} \geq n_{11} - \frac{I}{2}\right); \quad \hat{E}_{11} = \frac{n_{1\cdot}n_{\cdot 1}}{n}$$

En la Tabla 3 se observa cómo en los servicios “Limpieza, higiene, baños públicos”, “Carteles y señalización” y “Bancos/Cajeros automáticos” la decisión tomada respecto al rechazo o no de la versión positiva de la Teoría de la Asimilación en el caso del turismo cultural toledano depende de si al estadístico Chi-cuadrado se le incorpora la corrección propuesta en este trabajo, la de Yates, cualquiera de las dos de Cochran o la de Mantel. También puede apreciarse la gran diferencia en p-valor en el caso de “Entretenimiento y actividades culturales recreativas”.

Téngase en cuenta que en el caso que nos ocupa el tamaño muestral es ciertamente elevado. En caso de haber sido más pequeño el número de decisiones cambiantes hubiese sido superior. Y es que, a menor tamaño muestral, más relevante resulta la decisión sobre qué corrección de continuidad incluir en el estadístico Chi-cuadrado.

Tabla 3. Comparación de p-valores utilizando diferentes correcciones de continuidad.

Atributo	$n_{11} > \hat{E}_{11}$ (todos los casos)					Confirmación de la Teoría de la Asimilación (versión positiva)
	Montero	Yates	Cochran 1	Cochran 2	Mantel	
Entretenimiento y actividades culturales recreativas	0,1511	0,1263	0,1112	0,1261	0,0949	Montero NO Yates NO Cochran 1 NO Cochran 2 NO Mantel NO
Limpieza, higiene, baños públicos	0,0324*	0,0574	0,0528	0,0535	0,0562	Montero SI Yates NO Cochran 1 NO Cochran 2 NO Mantel NO
Carteles y señalización	0,0543	0,0391*	0,0422*	0,0471*	0,0449*	Montero SI Yates NO Cochran 1 NO Cochran 2 NO Mantel NO
Bancos/Cajeros automáticos	0,0314*	0,0572	0,0508	0,0533	0,0472*	Montero SI Yates NO Cochran 1 NO Cochran 2 NO Mantel NO

Fuente: elaboración propia.

* Significativo al 5%.

5. CONCLUSIONES.

La utilización de correcciones de continuidad en el estadístico Chi-cuadrado, a la hora de contrastar la hipótesis de independencia entre los niveles de dos o más factores, es una de las cuestiones clásicas de la estadística cualitativa sobre la que todavía en la actualidad se sigue investigando con profusión. Podría decirse que todavía es una de las cuestiones abiertas del ámbito de la estadística cualitativa.

Sin embargo, en este trabajo se propone una corrección de continuidad asimétrica que proporciona notables resultados en la denominada aproximación a la probabilidad exacta, o en otros términos, al p -valor. Y no solo es que proporcione buenos resultados sino que mejora notablemente las correcciones ya clásicas de Yates, Cochran y Mantel, entre otras.

En el ámbito del turismo cultural, una de las cuestiones en discusión en la actualidad es la sobreestimación de la satisfacción mostrada por el turista como consecuencia de la por nosotros denominada versión positiva de la Teoría de la Asimilación, esto es, como consecuencia de un conflicto psicológico que sufre el turista que le lleva a ajustar la satisfacción a las expectativas previas. Y es que en el marco del turismo cultural las expectativas suelen ser elevadas y suponer el cumplimiento de la versión positiva de la Teoría de la Asimilación implica una cierta sobreestimación de la satisfacción global de los turistas.

Sobre la base de un amplio cuestionario llevado a cabo en el casco histórico de la Ciudad de Toledo, se ha demostrado cómo en varios de los servicios considerados la decisión sobre si la Teoría de la Asimilación, en su versión positiva, opera en los visitantes de Toledo o no depende de la corrección de continuidad que se incluya en el estadístico de contraste, lo cual pone de manifiesto la importancia de la elección de ésta; en otros términos, la no inclusión de correcciones de continuidad asimétricas puede dar lugar a conclusiones erróneas, por lo cual cobra especial relevancia la corrección aquí expuesta..

Además, la relevancia de la elección de la corrección de continuidad es tanto mayor cuanto menor sea el número de cuestionarios disponible, circunstancia que realza el valor de la corrección asimétrica que se propone en este trabajo.

6. REFERENCIAS BIBLIOGRÁFICAS

- COCHRAN, W. G. (1942). "The 2x2 Correction for Continuity. Iowa State College" *Journal of Science*, 16, pp. 421-436.
- CONOVER, W.J. (1974). "Some reasons for not using the Yates continuity correction on 2x2 contingency tables". *Journal of the American Statistical Association*, 69, 346, pp. 374-376.
- CONOVER, W.J. (1968). "Uses and abuses of the continuity correction". *Biometrics*, 24, 1028.
- ESTEBAN, A., MARTÍN- CONSUEGRA, D., MOLINA, A., DÍAZ, E. (2005). "Turismo y consumo: El caso de Toledo". Centro de Estudios de Consumo de la Universidad de Castilla- La Mancha y de la Junta de Comunidades de Castilla- La Mancha. Working paper. Toledo.
- FLEISS, J.L. (1973). *Statistical Methods for Rates and Proportions*. New York: Wiley.
- HABER, M. (1980). "A Comparison of Some Continuity Corrections for the Chi-Squared Test on 2x2 Tables". *Journal of the American Statistical Association*, 75,371, pp. 510-515.
- HAVILAND, M.G. (2007). "Yates's correction for continuity and the analysis of 2 × 2 contingency tables". *Statistics in Medicine*, 9, 4, pp. 363-367.
- MARTÍN, A., TAPIA, J.M., SILVA-MATO, A., SÁNCHEZ, M. J. (2005). "On the Validity Condition of the Chi-Squared Test in 2x2 Tables". *Test*, 14, 1, pp.99-128.
- MANTEL, N. (1974). "Some Reasons for Not Using the Yates Continuity Correction on 2x2 Contingency Tables - Comment and a Sugestión". *Journal of the American Statistical Association*, 69, pp.378-380.
- MANTEL, N. (1976). "The Continuity Correction". *The American Statistician*, 30, pp.103-104.

- MANTEL, N., GREENHOUSE, S. (1968). "What is the Continuity Correction?". *The American Statistician*, 22, 5, pp.27-30.
- MONTERO LORENZO, J.M. (2002). "Una propuesta de corrección de continuidad asimétrica para tablas de contingencia (2x2) con totales marginales fijos". *Estadística Española*, 44, 149, pp. 29-46.
- OLIVER, R.L. (1997). *Satisfaction: A Behavioral Perspective on the Consumer*. New York: Mc Graw- Hill.
- RUIZ-MAYA, L., MARTÍN, J., MONTERO, J.M., URIZ, P. (1990). *Metodología Estadística para el análisis de datos cualitativos*. Madrid: C.I.S.
- RUIZ-MAYA, L., MARTÍN, J., MONTERO, J.M., URIZ, P. (1995). *Análisis Estadístico de Encuestas: Datos Cualitativos*. Madrid: A.C.
- SHERIF, M. & HOVLAND, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Press.
- YATES, F. (1934). "Contingency Tables Involving Small Numbers and the X^2 Test". *Journal of the Royal Statistical Society*, Ser. B, Supp., 1, 217-235.