Este documento forma parte de un conjunto de informes técnicos que explican la investigación llevada a cabo por los especialistas en estadística de Minitab para desarrollar los métodos y las verificaciones de los datos que se utilizan en el Asistente de Minitab Statistical Software.

Pruebas de chi-cuadrada

Revisión general

En la práctica, los profesionales de la calidad a veces necesitan recoger datos categóricos para evaluar un proceso cuando no es posible o conveniente recoger datos continuos. Por ejemplo, un producto puede clasificarse en dos categorías, como defectuoso/no defectuoso, o en más de dos categorías, como excelente, bueno, suficiente e insuficiente. Otro ejemplo es un departamento de finanzas que da seguimiento al número de días que las facturas permanecen vencidas clasificándolas en categorías: 15 días o menos, de 16 a 30 días, de 31 a 45 días o 45 días o más. Como resultado, la variable de interés es el número de elementos que entran en cada categoría.

Debido a su versatilidad, las pruebas de chi-cuadrada se usan para muchas aplicaciones que involucran datos categóricos. En el Asistente, las pruebas de chi-cuadrada se utilizan para:

- Probar la bondad de ajuste de una distribución multinomial

  Esta prueba puede utilizarse para determinar si los datos siguen la misma distribución que en el pasado. La distribución se define como una distribución multinomial con un conjunto de porcentajes históricos, u objetivo, que definen el porcentaje de elementos que entran en cada categoría de resultados. La prueba de chi-cuadrada evalúa de manera conjunta si algún porcentaje difiere significativamente de su porcentaje histórico u objetivo respectivo.

- Probar la igualdad del % de defectuosos para más de 2 grupos

  Esta prueba puede utilizarse para determinar si existe alguna diferencia entre el porcentaje de defectuosos de grupos diferentes. Los grupos se diferencian por una característica de interés, como un producto producido por distintos operadores, por diferentes plantas o en diferentes momentos. La prueba de chi-cuadrada evalúa de manera conjunta si cualquier porcentaje de defectuosos difiere significativamente de cualquier otro porcentaje de defectuosos.

- Probar la asociación entre dos variables categóricas

  Esta prueba puede utilizarse para determinar si una variable categórica de resultado (Y) está relacionada o asociada con otra variable categórica predictora (X). La prueba de chi-cuadrada evalúa de manera conjunta si existe una asociación entre la variable de resultado y una variable predictora. En el Asistente, usted puede realizar una prueba de chi-cuadrada para asociación con una variable predictora (X) que contenga dos o más valores distintos (dos o más muestras).

Para obtener más detalles sobre el estadístico de la prueba de chi-cuadrada, consulte el Apéndice A.

Para los métodos que implican la comprobación de hipótesis, es una buena práctica asegurarse de que se cumplan los supuestos de la prueba, que la prueba tenga suficiente potencia y que cualquier aproximación utilizada para analizar los datos produzca resultados válidos. Para las pruebas de chi-cuadrada, los supuestos son inherentes a la recolección de datos y no los abordamos en las verificaciones de los datos.

Centramos nuestra atención en la potencia y en la validez de los métodos de aproximación. El Asistente utiliza estos métodos de aproximación para realizar las siguientes verificaciones de sus datos y presenta los resultados en la Tarjeta de informe:

- Tamaño de la muestra
- Validez de la prueba
- Validez de los intervalos

En este trabajo, investigamos cómo se relacionan en la práctica estas verificaciones de los datos con las pruebas de chi-cuadrada y describimos cómo establecimos las directrices para las verificaciones de los de datos en el Asistente.

Verificaciones de los datos

Tamaño de la muestra

Por lo general, el principal objetivo de realizar una prueba estadística de hipótesis es reunir evidencia para rechazar la hipótesis nula de que "no existe diferencia”. Si las muestras son demasiado pequeñas, la potencia de la prueba podría no ser adecuada para detectar una diferencia en el porcentaje de defectuosos que en realidad existe, lo que produce un error Tipo II. Por lo tanto, resulta crucial asegurarse de que los tamaños de las muestras sean lo suficientemente grandes como para detectar diferencias importantes desde el punto de vista práctico con alta probabilidad.

La verificación de los datos con respecto al tamaño de la muestra se basa en la potencia de la prueba. Este cálculo requiere que el usuario especifique una diferencia significativa entre un parámetro de población real y el valor nulo hipotético. Como era demasiado difícil determinar y expresar esta diferencia práctica para la bondad de ajuste de chi-cuadrada y las pruebas de chi-cuadrada para asociación, el Asistente solo verifica el tamaño de la muestra para la prueba de chi-cuadrada del % de defectuosos con dos o más muestras.

Objetivo

Si los datos no proporcionan suficiente evidencia en contra de la hipótesis nula, debemos determinar si los tamaños de las muestras son lo suficientemente grandes como para que la prueba detecte diferencias de interés desde el punto de vista práctico con alta probabilidad. Aunque el objetivo de planificar el tamaño de la muestra es asegurar que los tamaños de las muestras sean lo suficientemente grandes como para detectar diferencias importantes con alta probabilidad, las muestras no deben ser tan grandes como para que las diferencias insignificantes se vuelvan estadísticamente significativas con alta probabilidad.

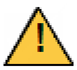
Método

El análisis de potencia y tamaño de la muestra se basa en las fórmulas que se muestran en el Apéndice B.

Resultados

Cuando los datos no ofrecen suficiente evidencia en contra de la hipótesis nula y usted no especifica una diferencia práctica, el Asistente calcula las diferencias prácticas que pueden detectarse con un 80% y un 90% de probabilidad con base en los tamaños de las muestras. Además, si el usuario especifica una diferencia práctica de interés en particular, el Asistente calcula los tamaños de las muestras que produzcan un 80% y un 90% de probabilidad de detección de esa diferencia.

Cuando se verifica la potencia y el tamaño de la muestra, la Tarjeta de informe del Asistente para la prueba de chi-cuadrada del % de defectuosos para más de dos muestras presenta los siguientes indicadores de estado:

Estado	Condición
	La prueba detecta una diferencia entre los % de defectuosos, así que la potencia no es un problema. O La potencia es suficiente. La prueba no detectó una diferencia entre los % de defectuosos, pero la muestra es lo suficientemente grande como para proporcionar al menos un 90% de probabilidad de detectar la diferencia dada.
	La potencia podría ser suficiente. La prueba no detectó una diferencia entre los % de defectuosos, pero la muestra es lo suficientemente grande como para ofrecer de un 80% a un 90% de probabilidad de detectar la diferencia dada. Se indica el tamaño de la muestra que se necesita para lograr una potencia de 90%.
	La potencia podría no ser suficiente. La prueba no detectó una diferencia entre los % de defectuosos y la muestra es lo suficientemente grande como para ofrecer de un 60% a un 80% de probabilidad de detectar la diferencia dada. Se indican los tamaños de las muestras que se necesitan para lograr una potencia de 80% y una potencia de 90%.
	La potencia no es suficiente (< 60%). La prueba no detectó una diferencia entre los % de defectuosos. Se indican los tamaños de las muestras que se necesitan para lograr una potencia de 80% y una potencia de 90%.
	La prueba no detectó una diferencia entre los % de defectuosos. Usted no especificó una diferencia práctica que detectar entre los % de defectuosos; por lo tanto, el informe indica las diferencias que se podrían detectar con una probabilidad de entre 80% y 90%, con base en los tamaños de sus muestras y el nivel de significancia.

Validez de la prueba

El estadístico de prueba χ^2 solo sigue aproximadamente una distribución de chi-cuadrada. La aproximación mejora con tamaños de muestra más grandes. En esta sección, evaluamos la aproximación utilizada para determinar el tamaño mínimo de la muestra que se necesita para obtener resultados exactos.

La aproximación de chi-cuadrada al estadístico de prueba se evalúa examinando el impacto de los pequeños conteos esperados de celdas sobre la tasa de error Tipo I (nivel de significancia). Usando el error Tipo I para evaluar la validez de la prueba, creamos una regla para asegurar que:

- La probabilidad de rechazar la hipótesis nula cuando es verdadera sea pequeña y esté cerca de la tasa de error Tipo I deseada.
- La cola de la distribución nula puede aproximarse razonablemente, lo que resulta importante para calcular el valor p con exactitud.

Usando un enfoque estándar, definimos un pequeño conteo esperado de celdas como una celda que tiene un conteo esperado de celdas menor que o igual a 5.

Desarrollamos dos modelos para definir las proporciones bajo la hipótesis nula: el modelo perturbado por las proporciones y el modelo de igual proporción. Para obtener más detalles, consulte el Apéndice C. Ambos modelos se utilizan en las simulaciones a las que nos referimos más adelante en este documento. Los modelos se usan para cada una de las pruebas de chi-cuadrada con una excepción: el modelo perturbado por las proporciones no se aplica a la prueba de chi-cuadrada del % de defectuosos para más de dos muestras.

La verificación de los datos con respecto a la validez de la prueba se aplica a todas las pruebas de chi-cuadrada disponibles en el Asistente. A continuación se describe cada una de las verificaciones de los datos.

Bondad de ajuste de chi-cuadrada

Objetivo

Evaluamos la aproximación de chi-cuadrada al estadístico de prueba investigando el impacto de la magnitud y la frecuencia de los pequeños conteos esperados de celdas sobre la tasa de error Tipo I.

Método

Se extrajeron muestras de tamaño n de una distribución multinomial con las proporciones descritas en el modelo perturbado por las proporciones o el modelo de igual proporción (consulte el Apéndice C). Para cada condición, realizamos 10,000 pruebas de bondad de ajuste de chi-cuadrada con un nivel de significancia objetivo de 0.05. Para cada prueba, calculamos el error Tipo I real como $\frac{Número\ de\ pruebas\ rechazadas}{Número\ de\ réplicas\ (10000)}$. Definimos un rango de tasas aceptables de error Tipo I de [0.03 – 0.07] y registramos el mínimo tamaño de muestra con una tasa de error Tipo I dentro de ese rango.

Resultados

Los resultados de la simulación revelaron que los conteos objetivo de celdas menores que 1.25 pueden conducir a valores p incorrectos cuando el porcentaje de pequeños conteos objetivo de celdas es menor que o igual a 50%. Además, los conteos objetivo de celdas menores que 2.5 puede conducir a valores p incorrectos cuando el porcentaje de pequeños conteos objetivo de celdas es mayor que 50%. Para obtener más detalles, consulte el Apéndice D.

Al verificar la validez de la prueba de bondad de ajuste de chi-cuadrada, la Tarjeta de informe del Asistente muestra los siguientes indicadores de estado:

Estado	Condición
	El mínimo conteo objetivo de celdas es mayor que o igual a 1.25 cuando el porcentaje de pequeños conteos objetivo de celdas es menor que o igual a 50% O El mínimo conteo objetivo de celdas es mayor que o igual a 2.5 cuando el porcentaje de pequeños conteos objetivo de celdas es mayor que 50%. Su muestra es lo suficientemente grande como para obtener suficientes conteos objetivo. El valor p de la prueba debe ser exacto.
	Si no se cumplen las condiciones anteriores.

Prueba de chi-cuadrada para asociación

Objetivo

Evaluamos la aproximación de chi-cuadrada al estadístico de prueba investigando el impacto de la magnitud y la frecuencia de los pequeños conteos esperados de celdas sobre la tasa de error Tipo I.

Método

Se extraen muestras de tamaño n_i de una distribución multinomial con las proporciones definidas en el modelo perturbado por las proporciones o en el modelo de igual proporción (consulte el Apéndice C). Por razones de simplicidad, elegimos $n_i = n \; \forall i$. Para cada condición, realizamos 10,000 pruebas de chi-cuadrada para asociación con un nivel de significancia de 0.05. Para cada prueba, calculamos la tasa real de error Tipo I como $\frac{Número\ de\ pruebas\ rechazadas}{Número\ de\ réplicas\ (10000)}$. Definimos un rango de tasas aceptables de error Tipo I de [0.03 – 0.07] y registramos el mínimo tamaño de muestra con una tasa de error Tipo I dentro de ese rango.

Resultados

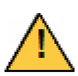
Encontramos que el mínimo conteo esperado de celdas depende de la cantidad de valores de X y del porcentaje de pequeños conteos esperados de celdas.

- Para el modelo perturbado por las proporciones, cuando el porcentaje de pequeños conteos esperados de celdas es menor que o igual a 50%, los mínimos conteos esperados de celdas son ≤ 2 y ≤ 1 para un número de valores de X igual a (2 ó 3) y (4, 5 ó 6), respectivamente. Además, cuando el porcentaje de pequeños conteos esperados de celdas es 50%, los mínimos conteos esperados de celdas son ≤ 3 y ≤ 1.5 para un número de valores de X igual a (2 ó 3) y (4, 5 ó 6), respectivamente.

- Para el modelo de igual proporción, el mínimo conteo esperado de celdas es ≤ 2 cuando el número de valores de X es igual a (2 ó 3) y el mínimo conteo esperado de celdas es ≤ 1.5 cuando el número de valores de X es igual a (4, 5 ó 6).

Para obtener más detalles, consulte el Apéndice E.

Al verificar la validez de la prueba de chi-cuadrada para asociación, la Tarjeta de informe del Asistente muestra los siguientes indicadores de estado:

Estado	Número de valores de la variable X	Condición
	2 ó 3	El mínimo conteo esperado de celdas es mayor que o igual a 2 cuando el porcentaje de pequeños conteos esperados de celdas (menor que o igual a 5) es menor que o igual a 50%. El mínimo conteo esperado de celdas es mayor que o igual a 3 cuando el porcentaje de pequeños conteos esperados de celdas (menor que o igual a 5) es mayor que 50%.
	4, 5 ó 6	El mínimo conteo esperado de celdas es mayor que o igual a 1 cuando el porcentaje de pequeños conteos esperados de celdas (menor que o igual a 5) es menor que o igual a 50%. El mínimo conteo esperado de celdas es mayor que o igual a 2 (redondeo de 1.5 a 2 por conveniencia) cuando el porcentaje de pequeños conteos esperados de celdas (menor que o igual a 5) es mayor que 50%.
	Todos los casos	Si no se cumplen las condiciones anteriores.

Prueba de chi-cuadrada de % de defectuosos para más de dos muestras

Objetivo

Evaluamos la aproximación de chi-cuadrada al estadístico de prueba investigando el impacto de la magnitud y la frecuencia de los pequeños conteos esperados de celdas sobre la tasa de error Tipo I.

Método

Definimos los modelos p = $p_i = p_j \ \forall i, j$, donde p = 0.001, 0.005, 0.01, 0.025 y 0.25. Se extrajeron muestras con un tamaño de n_i de una distribución binomial con los valores de p_i descritos anteriormente. Por razones de simplicidad, elegimos $n_i = n \ \forall i$. Para cada condición, realizamos 10,000 pruebas de chi-cuadrada de % de defectuosos con un nivel de significancia objetivo de 0.05. Para cada prueba, calculamos la tasa real de error Tipo I como $\frac{Número\ de\ pruebas\ rechazadas}{Número\ de\ réplicas\ (10000)}$. Definimos un rango de tasas aceptables de error Tipo I de [0.03 –

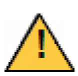
0.07] y registramos el mínimo tamaño de muestra con una tasa de error Tipo I dentro de ese rango.

Resultados

Cuando hay de 3 a 6 valores de X, un número mínimo esperado de defectuosos y no defectuosos mayor que o igual a 1.5 produce una tasa de error Tipo I para la prueba dentro del intervalo [0.03, 0.07]. Cuando hay de 7 a 12 valores de X, un número mínimo esperado de defectuosos y no defectuosos mayor que o igual a 1 produce una tasa de error Tipo I para la prueba dentro del intervalo [0.03, 0.07].

Para obtener más detalles, consulte el Apéndice F.

Cuando se verifica la validez de la prueba de chi-cuadrada del % de defectuosos para más de dos muestras, la Tarjeta de informe del Asistente presenta los siguientes indicadores de estado:

Estado	Número de valores de X	Condición
	De 3 a 6	El número mínimo esperado de defectuosos y no defectuosos es mayor que o igual a 1.5.
	De 7 a 12	El número mínimo esperado de defectuosos y no defectuosos es mayor que o igual a 1.
	Todos los casos	Si no se cumplen las condiciones anteriores.

Validez de los intervalos

Los intervalos de comparación en la prueba de chi-cuadrada del % de defectuosos para más de dos muestras y la prueba de bondad de ajuste de chi-cuadrada se basan en la aproximación a la normal. Además, los intervalos de confianza individuales en la prueba de bondad de ajuste de chi-cuadrada se basan en la aproximación a la normal. En esta sección, evaluamos la validez de la aproximación a la normal. De acuerdo con la regla general que se encuentra en la mayoría de los libros de texto de Estadística, el intervalo de confianza aproximado es exacto si los conteos observados son por lo menos 5.

La verificación de los datos con respecto a la Validez de los intervalos se aplica a la prueba de chi-cuadrada del % de defectuosos para más de dos muestras y la prueba de bondad de ajuste de chi-cuadrada.

Prueba de chi-cuadrada de % de defectuosos para más de dos muestras

Objetivo

Queríamos evaluar la regla general para el número mínimo de defectuosos y no defectuosos observados en cada muestra para asegurar que los intervalos de confianza aproximados sean exactos.

Método

Primero definimos los intervalos que se utilizan en la gráfica de comparación. Las cotas de los intervalos se definen de tal modo que con una tasa de error general de aproximadamente ∝, cualquier intervalo que no se superponga indica % de defectuosos de la población que son diferentes. Para ver las fórmulas utilizadas, consulte el Apéndice G.

Los intervalos de comparación se basan en intervalos de confianza de comparación en parejas. Para obtener más detalles, consulte la sección Intervalos de comparación en el Informe técnico del Asistente sobre el ANOVA de un solo factor. Utilizamos un intervalo de confianza de aproximación a la normal para cada par ($p_i - p_j$) y luego usamos un procedimiento de comparación múltiple de Bonferroni para controlar la tasa de error general por experimento. Por lo tanto, solo necesitamos evaluar la validez de uno de los intervalos en el procedimiento de comparación en parejas para entender el efecto de la aproximación a la normal sobre los intervalos de comparación.

Resultados

Para evaluar la validez de la aproximación a la normal, solo tenemos que examinar el efecto que tiene la aproximación sobre un intervalo para la diferencia entre los % de defectuosos. Por lo tanto, simplemente podemos usar la regla general creada para el caso del % de defectuosos para 2 muestras. Para obtener más detalles, consulte la sección Métodos de la prueba del % de defectuosos para 2 muestras en el Informe técnico del Asistente sobre la prueba del % de defectuosos para 2 muestras. Los resultados de la simulación en la prueba del % de defectuosos para 2 muestras indican que por lo general la exactitud del intervalo de confianza aproximado para la diferencia entre los % de defectuosos es fiable cuando las muestras son lo suficientemente grandes, es decir, cuando el número observado de defectuosos y el número observado de no defectuosos en cada muestra es por lo menos 5.

Cuando se verifica la validez de los intervalos para la prueba de chi-cuadrada del % de defectuosos para más de dos muestras, la Tarjeta de informe del Asistente presenta los siguientes indicadores de estado:

Estado	Condición
	Todas las muestras tienen al menos 5 defectuosos y 5 no defectuosos. Los intervalos de comparación deben ser exactos.
	Si no se cumple la condición anterior.

Bondad de ajuste de chi-cuadrada

Objetivo

Queríamos evaluar la regla general para el número mínimo de defectuosos y no defectuosos observados en cada muestra para asegurar que los intervalos de confianza aproximados sean exactos.

Método

La prueba de bondad de ajuste de chi-cuadrada del Asistente incluye intervalos de confianza de comparación e individuales. Podemos usar los intervalos estándar de aproximación a la normal para las proporciones y corregir para múltiples intervalos usando la corrección de Bonferroni (Goodman, 1965). Por lo tanto, los intervalos simultáneos de Bonferroni se calculan de la siguiente manera:

$$p_{iInferior} = p_i - Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1 - \mathrm{p_i})}{\mathrm{N}}}$$

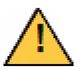
$$p_{iSuperior} = p_i + Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1 - \mathrm{p_i})}{\mathrm{N}}}$$

Las cotas de los intervalos se definen de tal modo que con una tasa de error general de aproximadamente ∝, cualquier intervalo que no contenga el valor de proporción objetivo indica que la proporción real es diferente de su proporción objetivo correspondiente. Los intervalos individuales utilizan la misma forma que los intervalos de Bonferroni, pero no realizan la corrección para intervalos múltiples usando $Z_{\alpha/2}$.

Resultados

Los dos métodos descritos anteriormente siguen una metodología similar a la que se define en la prueba del % de defectuosos para 2 muestras del Asistente. Por lo tanto, podemos usar reglas similares para la validez de la aproximación a la normal que se crearon para esa prueba. Para obtener más detalles, consulte la sección Métodos de la prueba del % de defectuosos para 2 muestras en el Informe técnico del Asistente sobre la prueba del % de defectuosos para 2 muestras. En ese documento, llegamos a la conclusión de que los intervalos de comparación y los intervalos de confianza individuales podrían no ser exactos cuando los conteos de muestras son menos que 5.

Al verificar la validez de los intervalos para la prueba de bondad de ajuste de chi-cuadrada, la Tarjeta de informe del Asistente muestra los siguientes indicadores de estado:

Estado	**Condición**
	Todos los conteos de muestra son de por lo menos 5. Los intervalos deberían ser exactos.
	Hay conteos de muestras menores que 5.

Referencias

Agresti, A. (1996). An introduction to categorical data analysis. New York, NY: Wiley.

Read, T. y Cressie, N. (1988). Goodness-of-fit statistics for discrete multivariate data. New York, NY: Springer-Verlag.

Fienberg, S. (1980). The analysis of cross-classified categorical data. Cambridge, MA: MIT Press.

Goodman, L. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics, 7*, 247-254.

Apéndice A: Estadístico de la prueba de chi-cuadrada

El Asistente usa un estadístico de la prueba de chi-cuadrada con la siguiente forma:

$$x^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

donde

O_{ij} = conteos observados, como se define en la siguiente tabla:

Caso	$\boldsymbol{O_{ij}}$
Probar la bondad de ajuste de una distribución multinomial	El conteo observado para el resultado de $i^{ésimo}$ se define como O_{i1}.
Probar la igualdad de más de 2 % de defectuosos	En número observado de elementos defectuosos y elementos no defectuosos para la muestra de $i^{ésimo}$ se define como O_{i1} y O_{i2} , respectivamente.
Probar la asociación entre dos variables categóricas	Los conteos observados para el valor $i^{ésimo}$ de la variable X y el valor $j^{ésimo}$ de la variable Y se definen como O_{ij}.

E_{ij} = Conteo esperado como se define en la siguiente tabla:

Caso	$\boldsymbol{E_{ij}}$
Probar la bondad de ajuste de una distribución multinomial	$E_{i1} = np_i$ $i = 1, \ldots, k$ (k = número de resultados) n = tamaño de la muestra p_i = proporciones históricas $\sum_i p_i = 1$
Probar la igualdad de más de 2 % de defectuosos	$E_{i1} = n_i p$ (para defectuosos) $E_{i2} = n_i(1 - p)$ (para no defectuosos) $i = 1, \ldots, k$ (k = número de muestras) n_i = $i^{ésimo}$ tamaño de la muestra p = proporción general de defectuosos

Caso	$\boldsymbol{E_{ij}}$
Probar la asociación entre dos variables categóricas	$E_{ij} = \frac{(n_{i.} n_{.j})}{n_{..}}$ $i = 1, \ldots, m$ (m = número de valores de X) $j = 1, \ldots, k$ (k = número de valores de Y) $n_{i.}$ = conteo total para el valor $i^{ésimo}$ de la variable X $n_{.j}$ = conteo total para el valor $j^{ésimo}$ de la variable Y $n_{..}$ = tamaño de muestra general

Apéndice B: Potencia de la prueba de chi-cuadrada del % de defectuosos para más de dos muestras

Utilizamos una distribución de chi-cuadrada no central para calcular la potencia de la prueba que $p_i = p_j = p \; \forall i, j$. El parámetro de no centralidad depende de n_i y $p_i \forall i$

donde

n_i = el tamaño de la muestra para la muestra de $i^{ésimo}$

Cada p_i representa una proporción alternativa (consulte la siguiente sección de este apéndice, Cálculo de proporciones alternativas) calculada con base en la diferencia de la proporción = δ.

Calculamos el parámetro de no centralidad de la distribución de chi-cuadrada como:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

donde

$$O_{i1 = n_i p_i}$$

$$O_{i2 = n_i (1 - p_i)}$$

y calculamos la potencia de la prueba como

$$\text{Prob}(X \geq x_{1-\alpha} \mid \chi^2)$$

donde

X = es una variable aleatoria de una distribución de chi-cuadrada no central con parámetro de no centralidad χ^2.

$x_{1-\alpha}$ = cdf inversa evaluada en $1 - \alpha$ para una distribución de chi-cuadrada central.

Cálculo de proporciones alternativas

Definimos las proporciones alternativas de la siguiente manera:

$$p_i = p_c + \frac{n_j}{n_i + n_j} \delta$$

$$p_j = p_c - \frac{n_i}{n_i + n_j} \delta$$

$$p_m = p_c \forall m \neq i, j$$

$$0 < \delta < 1$$

donde

$$p_c = \frac{1}{N_T}\sum_{i=1}^{k} n_i \widehat{p_\iota}$$

$\widehat{p_\iota}$ = elementos defectuosos de la proporción de muestra para la muestra $\iota^{ésimo}$.

NT = número total de observaciones.

n_i = tamaño de la muestra para la muestra $\iota^{ésimo}$.

Para algunas diferencias δ , $p_i > 1$ o $p_j < 0$. Por lo tanto, creamos las siguientes reglas:

Si p_j < 0

$$p_i = \delta$$
$$p_j = 0$$
$$p_m = \frac{\delta}{2}\ \forall m \neq i, j$$

Si p_i > 1

$$p_i = 1$$
$$p_j = 1 - \delta$$
$$p_m = 1 - \frac{\delta}{2}\ \forall m \neq i, j$$

Al usar los dos valores más pequeños de $n_{i,}$ se obtiene la potencia mínima, mientras que al usar los dos valores más grandes de n_i, se obtiene la potencia máxima.

Apéndice C: Modelo perturbado por las proporciones y modelo de igual proporción

Modelo perturbado por las proporciones

Siguiendo a Read y Cressie (1988), definimos el conjunto de proporciones bajo la hipótesis nula de la siguiente manera:

Elegimos δ cerca de k - 1 (donde k = número de proporciones para cada muestra) y definimos un conjunto de p_i pequeños como

$p_i = \frac{(1-\frac{\delta}{k-1})}{k}$ para $i = 1, \dots, r$

y el p_i restante como

$p_i = \frac{(1-\sum_{i=1}^{r} p_i)}{(k-r)}$ para $i = r+1, \dots, k$

Los valores que usamos para δ en las simulaciones se especifican en la tabla 1.

Tabla 1 δ utilizado en las simulaciones con los p_i pequeños resultantes

k	δ	$p_{i=1,\dots,r}$
3	1.95	0.008
4	2.95	0.004
5	3.90	0.005
6	4.90	0.003

Para cada k, variamos r = 1,..., k – 1 para cambiar el tamaño del conjunto de p_i's. pequeños. Por ejemplo, para k = 3, obtuvimos los siguientes dos modelos descritos en la tabla 2.

Tabla 2 Los valores de p_i para k = 3 usando el modelo perturbado por las las proporciones

r	p1	p2	p3
1	0.008	0.496	0.496
2	0.008	0.008	0.984

Modelo de igual proporción

Para obtener un modelo donde el 100% de los conteos esperados de celdas sean pequeños, utilizamos un modelo de igual proporción definido por

$$p_i = \frac{1}{k} \forall\, i$$

Cuando se usa este modelo, con un tamaño de muestra muy pequeño, todos los conteos esperados de celdas se consideran pequeños. Con un modelo de igual proporción, es necesario que los tamaños de las muestras sean muy pequeños para lograr un pequeño conteo esperado de celdas, lo que probablemente no ocurrirá en la práctica.

Apéndice D: Validez de la prueba para la bondad de ajuste de chi-cuadrada

Para el modelo perturbado por las proporciones, graficamos el mínimo conteo esperado de celdas necesario para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] en función de los pequeños conteos esperados de celdas, como se muestra en la figura 1.

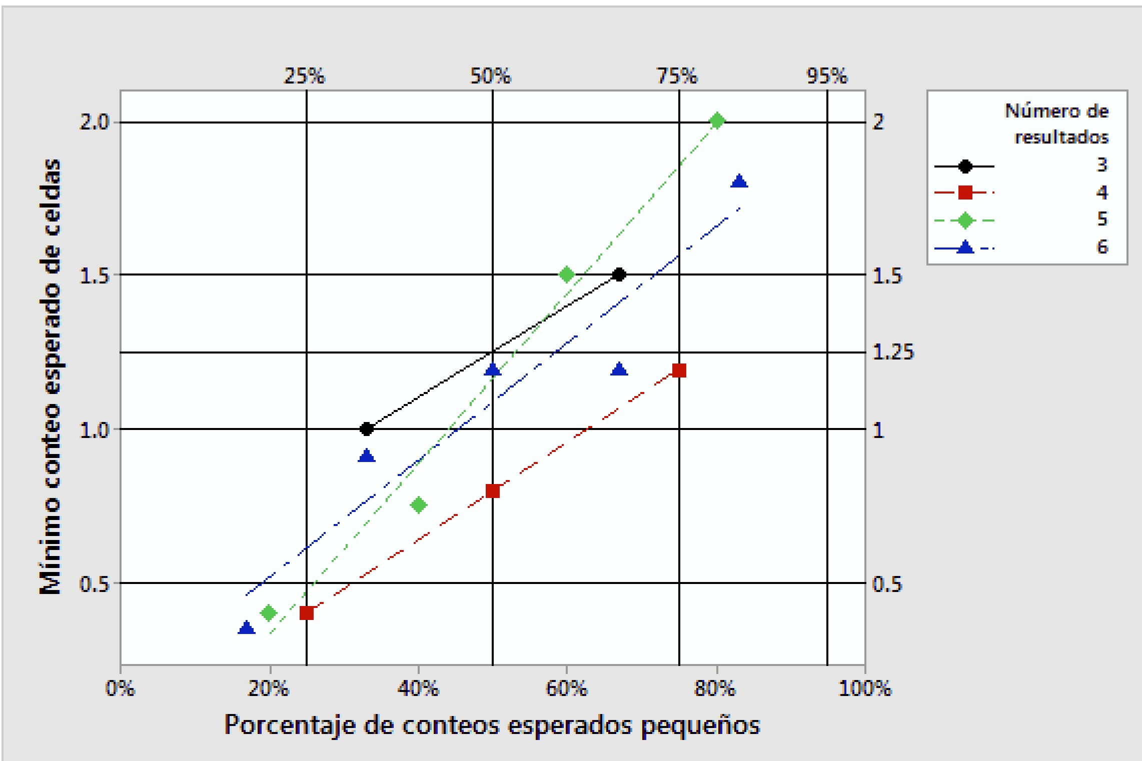


Figura 1 Mínimos conteos esperados de celdas necesarios para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] vs. el porcentaje de pequeños conteos esperados de celdas.

En la figura 1, cuando el porcentaje de pequeños conteos esperados de celdas es menor que 50%, los mínimos conteos esperados de celdas son menores que o iguales a 1.25. Todos los mínimos conteos esperados de celdas son menores que o iguales a 2. Con base en estos resultados de la simulación, las reglas que usamos en la Tarjeta de informe del Asistente son conservadoras.

Posteriormente, realizamos la misma simulación utilizando el modelo de igual proporción para definir la distribución nula. En la tabla 4 se resumen los resultados de la simulación utilizando un modelo de igual proporción.

Tabla 4 Mínimo conteo esperado de celdas para obtener una tasa de error Tipo I dentro del intervalo [0.03, 0.07]

k	Mínimo conteo esperado de celdas
3	2.5
4	1.25
5	1
6	1.4

Como se indicó anteriormente, el modelo de igual proporción conduce a casos en los que el 100% de los conteos de celdas son pequeños. La tabla 4 muestra que todos los mínimos conteos esperados de celdas son menores que o iguales a 2.5, lo que apoya las reglas que usamos en la Tarjeta de informe del Asistente.

Apéndice E: Validez de la prueba para la prueba de chi-cuadrada para asociación

Para el modelo perturbado por las proporciones, graficamos el mínimo conteo esperado de celdas necesario para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] en función del porcentaje de pequeños conteos esperados de celdas para cada número de valores de X, como se muestra en la figura 2.

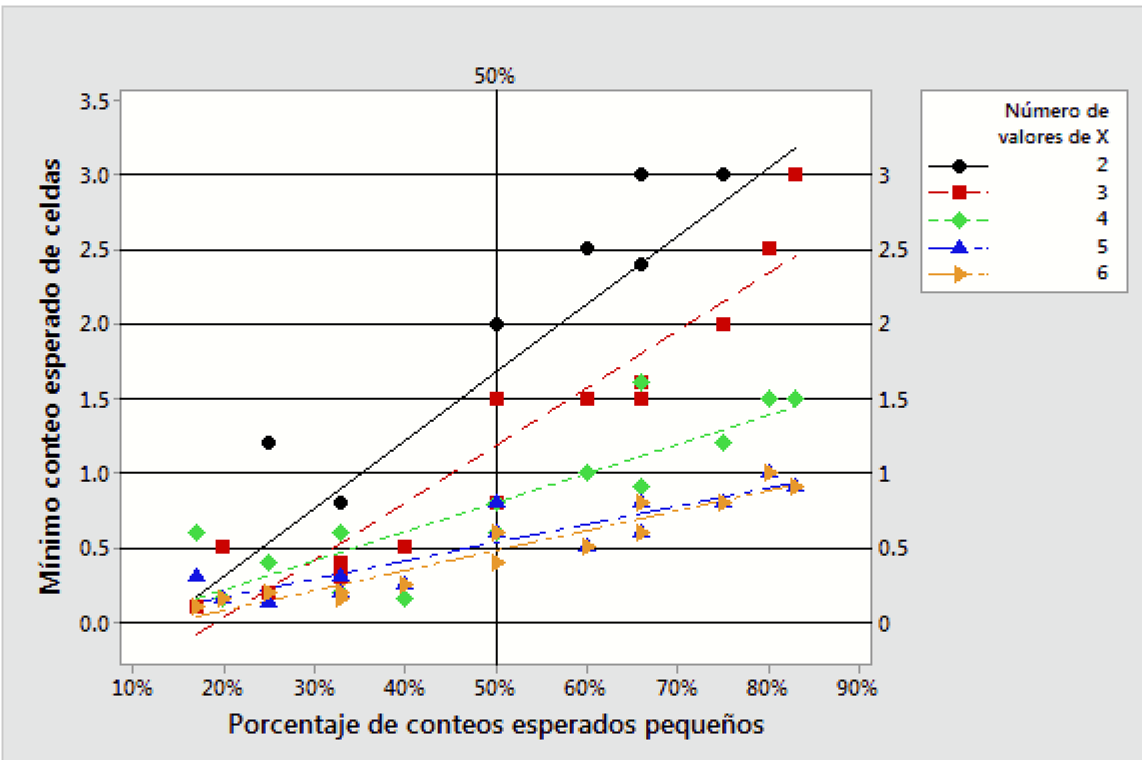


Figura 2 Mínimos conteos esperados de celdas necesarios para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] vs. el porcentaje de pequeños conteos esperados de celdas.

La figura 2 indica que el mínimo conteo esperado de celdas depende de la cantidad de valores de X y del porcentaje de pequeños conteos esperados de celdas.

La figura 2 indica que cuando el porcentaje de pequeños conteos esperados de celdas es ≤ 50%, los mínimos conteos esperados de celdas son ≤ 2 y ≤ 1 para un número de valores de X igual a 2 ó 3 y 4, 5 ó 6, respectivamente. Además, cuando el porcentaje de pequeños conteos esperados de celdas es > 50%, los mínimos conteos esperados de celdas son ≤ 3 y ≤ 1.5 para un número de valores de X igual a 2 ó 3 y 4, 5 ó 6, respectivamente.

Para el modelo de igual proporción, graficamos el mínimo conteo esperado de celdas en función del número de valores de X (m) y el número de valores de Y (k), como se muestra en la figura 3.

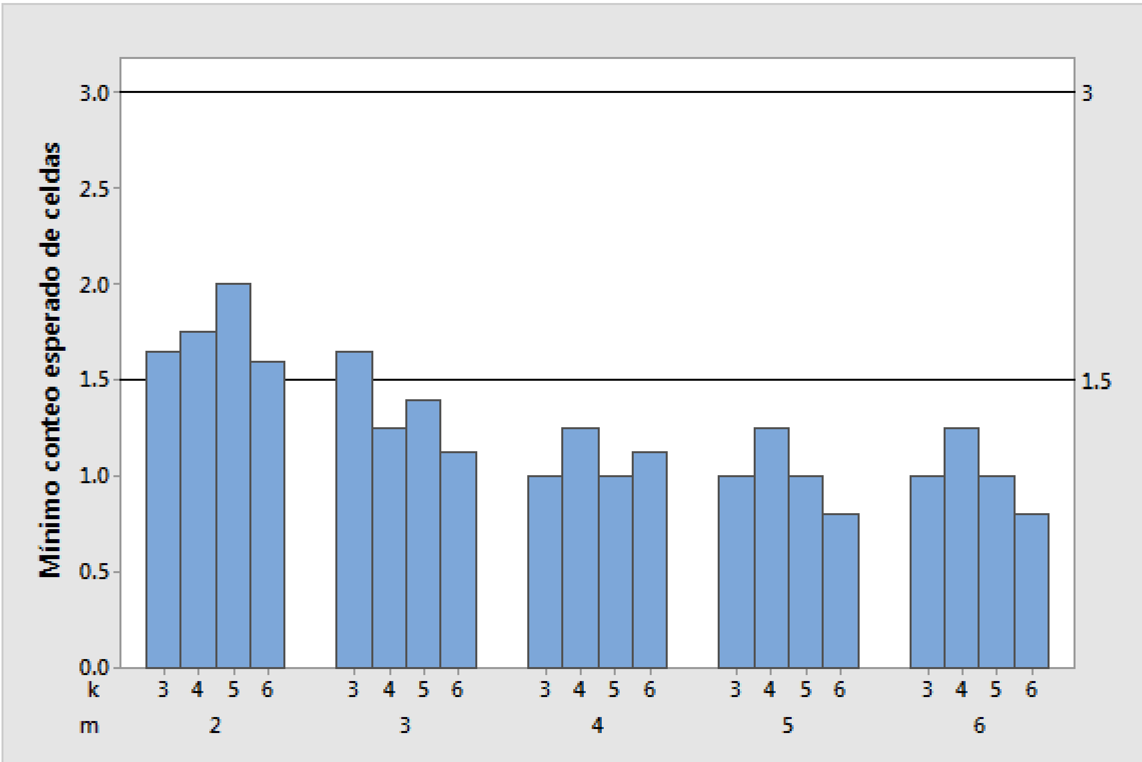


Figura 3 Mínimo conteo esperado de celdas necesario para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] vs. los valores de X (m) y los valores de Y (k).

La figura 3 indica que el mínimo conteo esperado de celdas es ≤ 2 cuando el número de valores de X es igual a 2 ó 3 y el mínimo conteo esperado de celdas es ≤ 1.5 cuando el número de valores de X es igual a 4, 5 ó 6. Con base en estos resultados de la simulación, las reglas utilizadas en la Tarjeta de informe del Asistente son conservadoras.

Apéndice F: Validez de la prueba de chi-cuadrada del % de defectuosos para más de dos muestras

Para cada p y cada m = 3, 4, 5,..., 12, graficamos el mínimo conteo esperado de celdas. Los resultados se muestran las figuras 4 y 5.

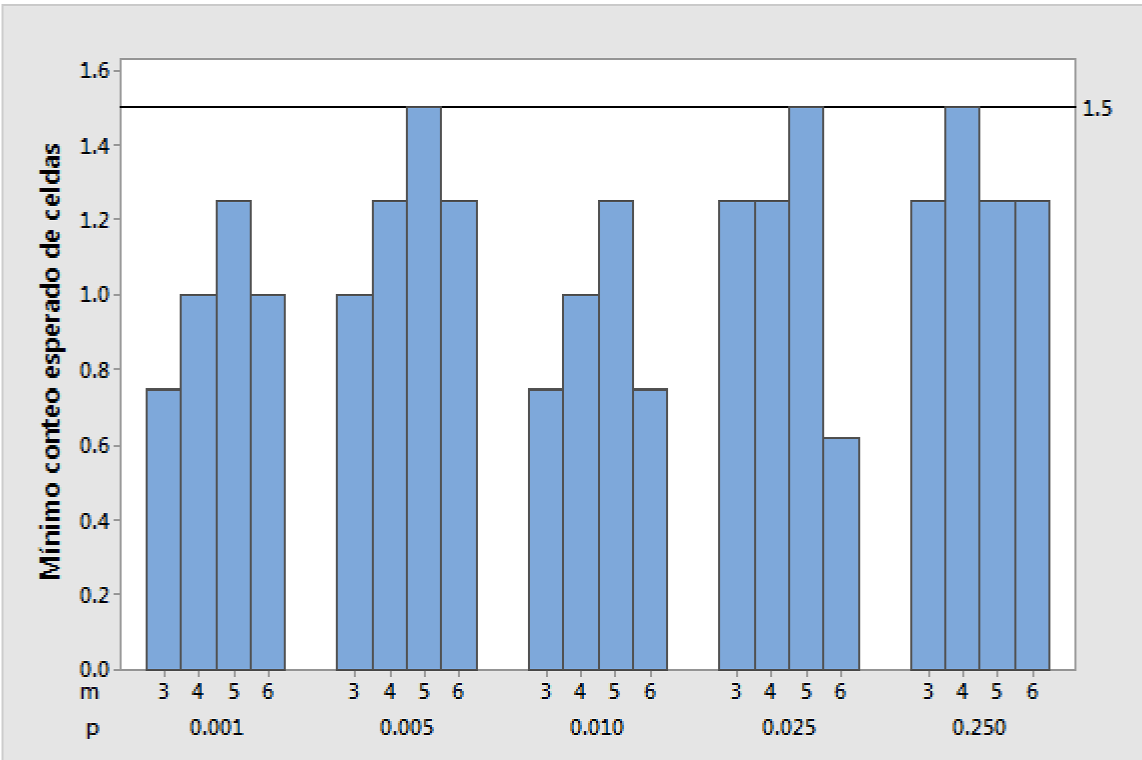


Figura 4 Mínimo conteo esperado de celdas necesario para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] vs. el número de valores de X (m = 3 hasta 6).

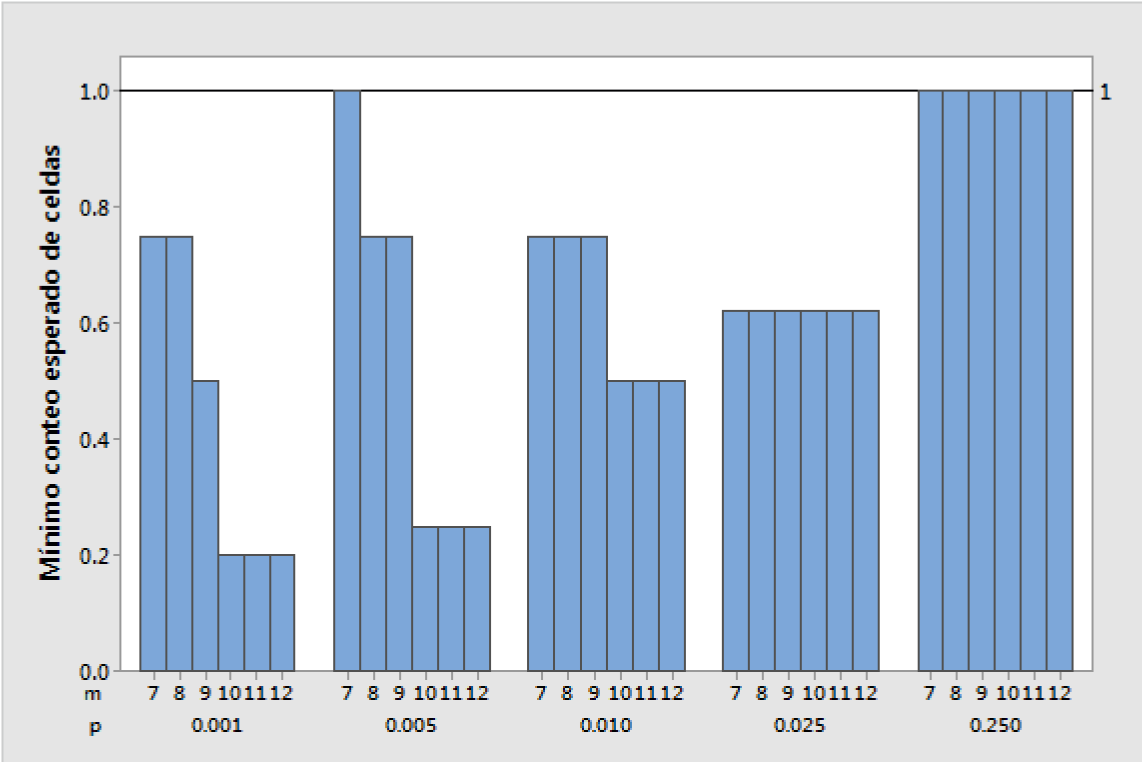


Figura 5 Mínimo conteo esperado de celdas necesario para lograr una tasa de error Tipo I dentro del intervalo [0.03, 0.07] vs. el número de valores de X (m = 7 hasta 12).

Cuando el número de valores de X es igual a 3, 4, 5 ó 6, un conteo esperado de celdas mayor que o igual a 1.5 produce una tasa de error Tipo I para la prueba dentro del intervalo [0.03, 0.07]. Cuando el número de valores de X es igual a 7, 8, 9,..., 12, un conteo esperado de celdas mayor que o igual a 1 produce una tasa de error Tipo I para la prueba dentro del intervalo [0.03, 0.07].

Apéndice G: Intervalos de comparación de la prueba de chi-cuadrada del % de defectuosos para más de dos muestras

Los límites superior e inferior de p_i se definen de la siguiente manera:

$$p_{iInferior} = p_i - Z_{\alpha/c} X_i$$

$$p_{iSuperior} = p_i + Z_{\alpha/c} X_i$$

donde

c = número de comparaciones = k (k - 1) /2

donde k es el número de muestras.

$Z_{\alpha/c}$ = (1 – percentil $\frac{\alpha}{2c}$) para una distribución normal con media = 0 y desviación estándar = 1

$$X_i = ((k - 1)\sum_{j \neq i} b_{ij} - \sum\sum_{1 \leq j < l \leq k} b_{jl}) / ((k - 1)(k - 2))$$

donde

$$b_{ij} = \sqrt{\frac{p_i(1 - p_i)}{n_i} + \frac{p_j(1 - p_j)}{n_j}}$$