

Análisis del error tipo I en las pruebas de bondad de ajuste e independencia utilizando el muestreo con parcelas de tamaño variable (Bitterlich)

Analysis of the Type I error in goodness of fit and independence tests, using variable radius plot sampling (Bitterlich)

MARIA ALEJANDRA QUINTERO M. Y MARIANO DURAN

Departamento de Manejo de Bosques, Escuela de Ingeniería Forestal, Facultad de Ciencias Forestales y Ambientales, Universidad de Los Andes, Av. Principal Chorro de Milla, Mérida, Venezuela.
E-mail: mariaq@ula.ve, mjdurann@ureach.com

SUMMARY

The Type I error for different goodness of fit and independence tests applied to data obtained from variable radius plot sampling (Bitterlich sampling) was studied. Five tests were evaluated: 1. the chi-square tests of Pearson, 2. Pearson considered with inclusion probabilities, 3. Wald, 4. Rao-Scott with first-order correction, and 5. Rao-Scott with second-order correction. Data obtained from both a plantation and from a natural forest were used. For different experimental conditions, a program that simulates variable radius plot sampling was used to estimate the Type I error of the goodness of fit and independence tests, applying the Monte Carlo method. The results of the investigation demonstrate that the chi-square tests of Pearson for goodness of fit and independence, which are commonly used techniques, register a distortion of the Type I error when compared with the nominal Type I error value ($\alpha=0.05$). The goodness of fit and independence test that gave the best results was the second-order correction Rao-Scott method..

Key words: goodness of fit, independence, chi-square test, Monte Carlo method.

RESUMEN

El error tipo I de diferentes pruebas de bondad de ajuste e independencia es estudiado para datos obtenidos mediante el muestreo con parcelas de tamaño variable. Se analizaron las pruebas chi-cuadrado de Pearson, Pearson ponderada por las probabilidades de inclusión, Wald, Rao-Scott con corrección de primer y segundo orden. Se utilizaron datos de una plantación y de un bosque natural, y mediante un programa que simula el muestreo con parcelas de tamaño variable usando el método de Monte Carlo, se estimó el error tipo I de las pruebas de bondad de ajuste e independencia, para distintas condiciones experimentales. Los resultados de la investigación demuestran que las pruebas chi-cuadrado de Pearson de bondad de ajuste e independencia, técnicas comúnmente usadas, registran una distorsión del error tipo I con respecto al valor nominal ($\alpha=0,05$). Las pruebas de bondad de ajuste e independencia que mostraron el mejor comportamiento son las de Rao-Scott con corrección de segundo orden.

Palabras clave: bondad de ajuste, independencia, parcelas de tamaño variable, pruebas chi-cuadrado, método de Monte Carlo.

INTRODUCCION

Las pruebas de bondad de ajuste e independencia son ampliamente utilizadas en diversas áreas de la ciencia para llevar a cabo análisis de datos categóricos. Una prueba de bondad de ajuste permite docimar la hipótesis de que una variable aleatoria sigue cierta distribución de probabilidad y se utiliza en situaciones donde se requiere comparar una distribución observada con una teórica o hipotética, compararla con datos históricos o con la distribución conocida de otra población (1). Por otra parte, una prueba de independencia permite determinar si existe asociación estadística entre dos variables categóricas.

Las pruebas más utilizadas para estudiar la independencia y la bondad de ajuste son las chi-cuadrado de Pearson, debido a su fácil aplicación y a que se encuentran en todos los paquetes estadísticos. Estas pruebas asumen que todas las observaciones son independientes y que están igualmente distribuidas, supuestos que sólo se satisfacen para un muestreo aleatorio simple con reposición y se cumplen aproximadamente en una muestra aleatoria simple sin reposición para una fracción de muestreo pequeña (2).

Está bien documentado en la literatura, aunque no es conocido en la práctica, que las pruebas chi-cuadrado de Pearson dan resultados extremadamente erróneos cuando se aplican en datos obtenidos mediante diseños muestrales complejos, donde no se cumplen los supuestos, tales como el muestreo sistemático, estratificado, por conglomerados, probabilidad variable o cualquier otro muestreo probabilístico diferente al muestreo aleatorio simple (3, 4). Las implicaciones por el reporte de resultados falsos pueden ser muy costosas, dependiendo de la naturaleza del estudio que se esté realizando.

Es una práctica común ignorar la complejidad del diseño muestral y proceder como si las pruebas chi-cuadrado de Pearson se comportaran de la misma forma que lo hacen bajo un muestreo aleatorio simple (5). De acuerdo a este criterio, se emplea un paquete estadístico estándar para aplicar las pruebas chi-cuadrado, lo que conduce a niveles de significación bastante altos (6).

En años recientes se ha realizado una serie de investigaciones para desarrollar nuevos métodos que tomen en cuenta la complejidad de la mues-

tra. Algunos de estos métodos son la prueba de Wald (7), los ajustes a las pruebas chi-cuadrado de Pearson propuestos por Fay (3) y las pruebas desarrolladas por Rao y Scott (8-10). Otros autores también han sugerido pruebas alternativas y han estudiado el error tipo I y la potencia de las pruebas chi-cuadrado de Pearson bajo diseños muestrales que violan los supuestos, tales como el estratificado y por conglomerados (11, 14).

El presente trabajo estudia el comportamiento de las pruebas chi-cuadrado de Pearson de bondad de ajuste e independencia, cuando se aplican en datos obtenidos mediante el muestreo de parcelas de tamaño variable, también conocido como el muestreo de Bitterlich o muestreo puntual. En este diseño muestral se viola el supuesto de independencia de las observaciones, ya que una vez que se elige el punto donde se aplica el muestreo, los árboles forman conglomerados y, por tanto, están presentes las correlaciones espaciales. Además, no se satisface el supuesto de igual probabilidad de selección para todas las unidades muestrales, pues los árboles son seleccionados con probabilidad proporcional a su área basal.

Para determinar si la violación de los supuestos afecta la validez de las pruebas de bondad de ajuste e independencia de Pearson, la investigación se fundamenta en la estimación del error tipo I de estas pruebas, aplicadas en datos obtenidos mediante un muestreo de parcelas de tamaño variable. El error tipo I se define como la probabilidad de rechazar la hipótesis nula siendo verdadera. También se le conoce como nivel de significación (α), y es fijado por el investigador. En este trabajo se estableció un α igual a 0,05, el cual es comparado con el error tipo I estimado a partir de un proceso de simulación ($\hat{\alpha}$).

Igualmente, se evalúa el error tipo I de algunas pruebas de bondad de ajuste e independencia que toman en cuenta el diseño muestral; estas son la prueba chi-cuadrado de Pearson ponderada por las probabilidades de inclusión, la prueba de Wald y las pruebas de Rao-Scott con corrección de primer y segundo orden.

Los resultados de este trabajo proporcionan a los investigadores forestales una idea acerca de cuáles pruebas de bondad de ajuste e independencia deben o no usarse, cuando se utiliza el muestreo con parcelas de tamaño variable, para evitar llegar a conclusiones erróneas.

MATERIAL Y METODOS

PRUEBAS DE BONDAD DE AJUSTE

En una prueba estadística de este tipo la hipótesis nula es:

$$H_0: P_i = P_{io} \text{ para } i = 1, 2, \dots, k$$

Donde P_i es la proporción de individuos que pertenecen a la categoría i , P_{io} es la proporción teórica de la categoría i , y k es el número de categorías.

En las siguientes secciones se presentan las fórmulas de cálculo de las distintas pruebas de bondad de ajuste aplicadas en esta investigación.

Prueba chi-cuadrado de Pearson: se basa en la diferencia entre las frecuencias o proporciones observadas y las esperadas.

El estadístico de prueba es el siguiente:

$$X_p^2 = n \sum_{i=1}^k \frac{(\hat{p}_i - P_{io})^2}{P_{io}} \quad [1]$$

Donde \hat{p}_i se obtiene al dividir número de individuos de la categoría i observados entre el total de individuos ($\hat{p}_i = n_i / n$).

El estadístico X_p^2 sigue una distribución chi-cuadrado con $k-1$ grados de libertad.

Prueba chi-cuadrado de Pearson ponderada por las probabilidades de inclusión: utiliza el mismo estadístico que la prueba chi-cuadrado de Pearson, con la diferencia de que \hat{p}_i se calcula usando ponderaciones que consideran las probabilidades de inclusión.

El estimador de p_i es el siguiente:

$$\hat{p}_i = \frac{\sum_{j=1}^n W_j Y_{ij}}{\sum_{j=1}^n W_j} \text{ con } W_j = 1 / j \quad [2]$$

Donde j es la probabilidad de que el individuo j sea seleccionado en la muestra (probabilidad de inclusión del individuo j) y Y_{ij} es una variable indicadora que vale 1 si el individuo j pertenece a la categoría i y 0 en otro caso.

Prueba de Wald: toma en consideración la correlación existente dentro de conglomerados, lo que hace que sea un procedimiento válido para probar la bondad del ajuste en algunos diseños muestrales complejos.

El estadístico de Wald para probar la bondad de ajuste es:

$$X_w^2 = (\hat{\mathbf{p}} - \mathbf{p}_0)' \hat{\mathbf{V}}^{-1} (\hat{\mathbf{p}} - \mathbf{p}_0) \quad [3]$$

Donde $\hat{\mathbf{p}}$ es el vector de proporciones estimadas $(\hat{p}_1, \dots, \hat{p}_{k-1})'$; $\hat{\mathbf{V}}$ es un estimador de la matriz de covarianza de $\hat{\mathbf{p}}$; y \mathbf{p}_0 es el vector de proporciones teóricas $(p_{o1}, \dots, p_{ok-1})'$.

El estadístico X_w^2 se distribuye aproximadamente como una chi-cuadrado con $k-1$ grados de libertad. El estimador $\hat{\mathbf{V}}$ puede ser obtenido por el método de linearización, replicación balanceada o Jackknife (9,15).

Prueba de Rao-Scott con corrección de primer orden: esta prueba propone un ajuste al estadístico chi-cuadrado de Pearson, siendo su propósito corregir la esperanza asintótica de este estadístico.

El estadístico de prueba es el siguiente:

$$X_{RSI}^2 = \frac{X_{pp}^2}{\hat{\delta}} \quad [4]$$

Donde X_{RSI}^2 es el estadístico chi-cuadrado Rao-Scott con corrección de primer orden; X_{pp}^2 es el estadístico chi-cuadrado de Pearson ponderado por las probabilidades de inclusión; y $\hat{\delta}$ es un estimador de la media $\bar{\delta}$ de los valores propios δ_j de la matriz de los efectos del diseño generalizados.

La ecuación para calcular $\bar{\delta}$ es:

$$\hat{\delta} = \sum_{i=1}^k \frac{n \cdot \hat{v}(\hat{p}_i)}{P_{oi} (k-1)} \quad [5]$$

Donde $\hat{v}(\hat{p}_i)$ es la varianza de los estimadores de proporciones ponderados por las probabilidades de inclusión.

El estadístico X_{RSI}^2 se distribuye aproximadamente como una chi-cuadrado con $k-1$ grados de libertad.

Prueba de Rao-Scott con corrección de segundo orden: corrige la esperanza y la varianza asintótica del estadístico chi-cuadrado de Pearson.

El estadístico con corrección de segundo orden es el siguiente:

$$X^2_{RS2} = \frac{X^2_{RS1}}{1 + \hat{a}^2} \quad [6]$$

Donde \hat{a} es un estimador del coeficiente de variación de los valores propios desconocidos δ_i , el cual está dado por la siguiente ecuación:

$$\hat{a}^2 = \sum_{i=1}^{k-1} \frac{\hat{\delta}_i^2}{(k-1)\hat{\delta}^2} - 1 \quad [7]$$

$\hat{\delta}$ se calcula mediante la ecuación 5 y $\sum_{i=1}^{k-1} \hat{\delta}_i^2$ de la siguiente manera:

$$\sum_{i=1}^{k-1} \hat{\delta}_i^2 = n^2 \sum_{i=1}^k \sum_{j=1}^k \frac{\hat{V}^2(\hat{p}_i, \hat{p}_j)}{p_{oi} \cdot p_{oj}} \quad [8]$$

siendo $\hat{V}^2(\hat{p}_i, \hat{p}_j)$ el elemento i,j de la matriz de covarianza de los estimadores de proporciones; p_{oi} y p_{oj} corresponden a las proporciones teóricas para las categorías i y j.

El estadístico X^2_{RS2} se distribuye aproximadamente como una chi-cuadrado con g grados de libertad. Los grados de libertad se calculan así:

$$g = \frac{k-1}{1 + \hat{a}^2} \quad [9]$$

PRUEBAS DE INDEPENDENCIA

Para analizar la independencia entre dos variables categóricas, la variable 1 con r categorías y la variable 2 con c categorías, los datos se organizan en una tabla de contingencia como la que se muestra a continuación:

Variable 1	Variable 2				Total
	1	2	c	
1	P ₁₁	P ₁₂	P _{1c}	P _{1.}
2	P ₂₁	P ₂₂	P _{2c}	P _{2.}
·	·	·	·	·	·
·	·	·	·	·	·
·	·	·	·	·	·
r	P _{r1}	P _{r2}	P _{rc}	P _{r.}
Total	P _{.1}	P _{.2}	P _{.c}	n

Donde P_{ij} es la proporción de individuos que pertenecen a la categoría i de la variable fila y a la categoría j de la variable columna; P_{i.} es $\sum_{j=1}^c P_{ij}$; y P_{.j} es $\sum_{i=1}^r P_{ij}$

La hipótesis nula de independencia es:

$$H_0: p_{ij} = p_{i.} \times p_{.j} \text{ para } i= 1, \dots, r \text{ } j= 1, \dots, c$$

Prueba chi-cuadrado de Pearson: utiliza un estadístico de prueba basado en la diferencia entre los valores observados y los esperados.

El estadístico chi-cuadrado de Pearson es:

$$X^2_p = n \sum_{i=1}^r \sum_{j=1}^c \frac{(\hat{p}_{ij} - \hat{p}_{i.} \cdot \hat{p}_{.j})^2}{\hat{p}_{i.} \cdot \hat{p}_{.j}} \quad [10]$$

Donde \hat{p}_{ij} se obtiene al dividir número de individuos observados en la celda i,j entre el total de individuos ($\hat{p}_i = n_{ij} / n$); $\hat{p}_{i.} = \sum_{j=1}^c \hat{p}_{ij}$ y $\hat{p}_{.j} = \sum_{i=1}^r \hat{p}_{ij}$.

El estadístico X^2_p se distribuye aproximadamente como una variable aleatoria chi-cuadrado con (r-1)(c-1) grados de libertad.

Prueba chi-cuadrado de Pearson ponderada por las probabilidades de inclusión: utiliza el mismo estadístico de la prueba chi-cuadrado de Pearson clásica, pero el estimador \hat{p}_{ij} está ponderado por las probabilidades de inclusión, tal como puede observarse en la ecuación 11.

$$\hat{p}_{ij} = \frac{\sum_{k=1}^n W_k Y_{ijk}}{\sum_{k=1}^n W_k} \quad \text{con } W_k = 1 / k \quad [11]$$

Donde \hat{p}_{ij} es la proporción estimada para el grupo o celda i,j ; $1/k$ es la probabilidad de que el individuo k sea seleccionado en la muestra (probabilidad de inclusión del individuo k); y Y_{ijk} es una variable indicadora que vale 1 si el individuo k pertenece a la celda i,j y 0 en otro caso.

Prueba de Wald: para aplicar esta prueba se reformula la hipótesis nula de independencia $H_0: p_{ij} = p_{i.} \times p_{.j}$, en una hipótesis equivalente $H_0: F_{ij} = p_{ij} - p_{i.} \times p_{.j} = 0$. Los valores F_{ij} se denominan diferencias residuales.

El estadístico de prueba o estadístico de Wald es el siguiente:

$$X^2_w = \hat{\mathbf{F}}' \hat{\mathbf{V}}_F^{-1} \hat{\mathbf{F}} \quad [12]$$

Donde $\hat{\mathbf{F}} = (\hat{F}_{11}, \dots, \hat{F}_{1,c-1}, \dots, \hat{F}_{r-1,1}, \dots, \hat{F}_{r-1,c-1})'$ es el vector de las diferencias residuales estimadas y $\hat{\mathbf{V}}_F$ estimador de la matriz de covarianza del vector $\hat{\mathbf{F}}$.

Las diferencias residuales estimadas son iguales a $\hat{F}_{ij} = \hat{p}_{ij} - \hat{p}_{i.} \times \hat{p}_{.j}$, y utilizan los estimadores de proporciones ponderados por las probabilidades de inclusión.

El estadístico X^2_w se distribuye aproximadamente como una chi-cuadrado con $(r-1)(c-1)$ grados de libertad. El estimador $\hat{\mathbf{V}}_F$ puede ser obtenido por el método de linearización, replicación balanceada o Jackknife, de igual manera que en la prueba de bondad de ajuste de Wald (9).

Prueba de Rao-Scott con corrección de primer orden: el estadístico que utiliza esta prueba de independencia es el mismo de la ecuación 4, pero $\hat{\delta}$ se calcula de acuerdo a la ecuación 13.

$$\hat{\delta} = \sum_{i=1}^r \sum_{j=1}^c \frac{n \hat{v}_{ij}(\hat{\mathbf{F}})}{(r-1)(c-1) \hat{p}_{i.} \times \hat{p}_{.j}} \quad [13]$$

Donde $\hat{v}_{ij}(\hat{\mathbf{F}})$ es el estimador de la varianza de \hat{F}_{ij} ; $\hat{p}_{i.}$ es el estimador de la proporción marginal

para la fila i ; y $\hat{p}_{.j}$ es el estimador de la proporción marginal para la columna j .

El estadístico de Rao-Scott con corrección de primer orden se distribuye aproximadamente como una chi-cuadrado con $(r-1)(c-1)$ grados de libertad.

Prueba de Rao-Scott con corrección de segundo orden: a continuación se especifican los pasos que deben seguirse para aplicar esta prueba de independencia (16).

1. Construir dos vectores denominados Y y P, los cuales tienen la siguiente forma:

$$Y = \sqrt{n} \begin{bmatrix} \hat{p}_{11} - \hat{p}_{1.} \times \hat{p}_{.1} \\ \hat{p}_{12} - \hat{p}_{1.} \times \hat{p}_{.2} \\ \vdots \\ \hat{p}_{1c} - \hat{p}_{1.} \times \hat{p}_{.c} \\ \hat{p}_{21} - \hat{p}_{2.} \times \hat{p}_{.1} \\ \hat{p}_{22} - \hat{p}_{2.} \times \hat{p}_{.2} \\ \vdots \\ \hat{p}_{2c} - \hat{p}_{2.} \times \hat{p}_{.c} \\ \vdots \\ \hat{p}_{r1} - \hat{p}_{r.} \times \hat{p}_{.1} \\ \hat{p}_{r2} - \hat{p}_{r.} \times \hat{p}_{.2} \\ \vdots \\ \hat{p}_{rc} - \hat{p}_{r.} \times \hat{p}_{.c} \end{bmatrix} \quad P = \begin{bmatrix} \hat{p}_{1.} \times \hat{p}_{.1} \\ \hat{p}_{1.} \times \hat{p}_{.2} \\ \vdots \\ \hat{p}_{1.} \times \hat{p}_{.c} \\ \hat{p}_{2.} \times \hat{p}_{.1} \\ \hat{p}_{2.} \times \hat{p}_{.2} \\ \vdots \\ \hat{p}_{2.} \times \hat{p}_{.c} \\ \vdots \\ \hat{p}_{r.} \times \hat{p}_{.1} \\ \hat{p}_{r.} \times \hat{p}_{.2} \\ \vdots \\ \hat{p}_{r.} \times \hat{p}_{.c} \end{bmatrix}$$

2. Calcular la matriz de covarianza del vector Y, llámese B.
3. Calcular los grados de libertad g mediante la siguiente fórmula:

$$g = \frac{\left(\sum_{i=1}^{rc} \frac{B_{ij}}{P_i} \right)^2}{\sum_{i=1}^{rc} \sum_{j=1}^{rc} \frac{B_{ij}^2}{P_i P_j}} \quad [14]$$

Siendo B_{ij} es elemento i,j de la matriz de covarianza del vector Y ; y P_i es el elemento i del vector P .

4. Calcular el estadístico de Rao-Scott usando la siguiente expresión:

$$X_{RS2}^2 = \frac{gX_{pp}^2}{\sum_{i=1}^r \frac{B_{ii}}{P_i}} \quad [15]$$

Donde X_{pp}^2 es el estadístico chi-cuadrado de Pearson ponderado por las probabilidades de inclusión.

El estadístico X_{RS2}^2 se distribuye aproximadamente como una variable aleatoria chi-cuadrado con g grados de libertad.

Para profundizar sobre los fundamentos teóricos de las pruebas de bondad de ajuste e independencia, se recomienda consultar Lehtonen y Pahkinen (15), Sarndal *et al.* (2) y los trabajos de Rao y Scott (8-10).

PROGRAMA DE SIMULACION

Se construyó un programa de simulación usando el lenguaje GAUSS versión 3.1.4. para obtener muestras mediante el método de parcelas de tamaño variable, aplicando el método de Monte Carlo. Usando las observaciones de una muestra, el programa puede aplicar todas las pruebas de bondad de ajuste e independencia: Pearson, Pearson ponderada, Wald, Rao-Scott con corrección de primer y segundo orden.

Para simular el muestreo con parcelas de tamaño variable, el programa requiere de una base de datos con la información de la población de árboles de donde se extraerá la muestra, la cual debe incluir la ubicación de cada árbol expresada como coordenadas cartesianas (x,y) , el DAP y los valores de las variables categóricas que se utilizarán en las pruebas de bondad de ajuste e independencia. También es necesario indicarle al programa las dimensiones del área a muestrear, el número de puntos muestrales y el ángulo de barrido.

A partir de estos datos, el programa obtiene una muestra simulando el procedimiento del muestreo con parcelas de tamaño variable; esto es, se seleccionan al azar varios puntos en el sitio

de estudio y desde cada uno, el observador, haciendo uso de un instrumento generador de ángulos (relascopio de Bitterlich), hace un giro de 360° en el sentido de las agujas del reloj e incluye en la muestra todos aquellos árboles cuyo DAP es intersectado por la proyección de las líneas que forman el ángulo de barrido (17-19).

Además, el programa aplica el método Mirage para corregir el llamado "efecto del borde" en los límites del área a muestrear. Dicho efecto se refiere al hecho de que un árbol ubicado cerca de los límites de la parcela puede tener una probabilidad de inclusión diferente a la de un árbol similar que esté situado hacia el interior de la parcela, originando un sesgo en los estimadores (19). El procedimiento que establece el método Mirage para corregir el efecto del borde en el muestreo con parcelas de tamaño variables es el siguiente (20): cada punto muestral establecido es proyectado ortogonalmente cruzando los bordes del área en estudio, originándose un punto de reflexión, llamado punto *mirage*. Este punto de reflexión y el punto muestral original deben ser equidistantes al borde sobre una línea perpendicular (conceptual) a éste. El muestreo se llevará a cabo haciendo el barrido con el relascopio de Bitterlich desde cada punto muestral y desde cada punto *mirage*. De esta forma, algunos árboles son incluidos en la muestra dos veces: una vez cuando se hace el barrido con el relascopio desde el punto muestral, y otra cuando se hace desde el punto *mirage*.

PROCEDIMIENTO PARA ESTIMAR EL ERROR TIPO I

Una vez construido el programa de simulación, el siguiente paso fue estimar el error tipo I de cada una de las pruebas de bondad de ajuste e independencia. El estimador utilizado es el siguiente:

$$\hat{\alpha} = \frac{A}{r} \quad [16]$$

Donde $\hat{\alpha}$ es el error tipo I estimado, A es el número de veces que se rechaza H_0 siendo verdadera y r es el número de replicaciones del programa de simulación (número de muestras).

Para estimar el error tipo I mediante la ecuación 16 es necesario establecer una hipótesis nula que sea verdadera en la población de árboles y

probar dicha hipótesis r veces, cada vez con una muestra diferente. Para ello, deben hacerse r corridas del programa de simulación, el cual se encarga de contar el número de veces que se rechaza la hipótesis nula y al final estima el error tipo I de las diferentes pruebas.

En todos los experimentos realizados se usó $r = 10.000$ para garantizar precisión en el estimador del error tipo I. Es fácil verificar que mientras mayor sea el número de replicaciones mayor será la precisión de la estimación (1).

Fue estimado el error tipo I de las pruebas de bondad de ajuste e independencia para diferentes condiciones experimentales, con el objetivo de comparar el $\hat{\alpha}$ con el valor nominal establecido ($\alpha = 0,05$) y determinar si hay distorsión.

Se utilizaron dos bases de datos. La primera de ellas correspondiente a una plantación de eucalipto, con una superficie de 4,85 hectáreas y 4.657 árboles. La segunda corresponde a un bosque natural, con una superficie de 1,62 hectáreas y 5.503 árboles. En ambas bases de datos se construyó la variable categoría diamétrica a partir de los valores del DAP, para lo cual se dividió el rango de valores observados de DAP en k intervalos de la misma amplitud y se numeraron de 1 a k . Para cada árbol se determinó en qué intervalo se ubica su DAP y se le asignó a la variable categoría diamétrica el valor correspondiente (1, 2, ..., k). De esta forma, se definieron tres variables categóricas, con $k=3$, $k=5$ y $k=10$, las cuales fueron utilizadas en las pruebas de bondad de ajuste e independencia. En la segunda base de datos se disponía una variable categórica adicional, pues cada árbol tenía asignado el valor 1, 2 ó 3, dependiendo de su especie forestal.

En el análisis del error tipo I de las pruebas de independencia fue necesario generar mediante simulación variables aleatorias con distribución binomial y multinomial, ya que las variables que se tenían en las bases de datos no eran independientes. De esta manera, se probó la independencia entre la variable categoría diamétrica ($k=3$) y una variable binomial, y entre la variable categoría diamétrica ($k=3$) y una variable multinomial con 3 categorías, usando las dos bases de datos y cambiando la cantidad de puntos muestrales iniciales. Las pruebas se hicieron para tablas de contingencia 3×2 y 3×3 por ser las más utilizadas.

RESULTADOS Y DISCUSION

En el cuadro 1 se presentan los valores del error tipo I estimado ($\hat{\alpha}$) de las diferentes pruebas de bondad de ajuste, después de ejecutar 10.000 repeticiones del programa de simulación, usando dos bases de datos, diferentes valores de k , y para 5, 10 y 15 puntos muestrales. Cabe señalar que en promedio se seleccionaron 10,06 árboles en cada punto muestral cuando se trabaja con la base de datos 1 y para un ángulo de barrido de $1,6205^\circ$; y se incluyen 12,5 árboles por punto muestral para la base de datos 2 y un ángulo de barrido igual a $1,1459^\circ$.

Puede observarse en el cuadro 1 que para un número considerable de situaciones se registran valores muy elevados de $\hat{\alpha}$, bastante diferentes del valor nominal ($\alpha = 0,05$). El error tipo I más alto se presenta en la prueba de bondad de ajuste chi-cuadrado de Pearson, siendo en todos los casos superior a 0,5. Esta prueba muestra un desempeño bastante deficiente.

Al usar la prueba chi-cuadrado de Pearson ponderada por las probabilidades de inclusión, el error tipo I estimado disminuye con respecto a la prueba chi-cuadrado de Pearson clásica, sin embargo, sigue siendo inaceptable, pues oscila entre 0,1810 y 0,4154.

La prueba de Wald registró en algunas situaciones un error tipo I estimado menor que el de la prueba de Pearson ponderada, en otros casos el error fue mayor. Se observa que la prueba de Wald tiene un error tipo I más alto cuando se trabaja con mayor número de categorías.

En todas las simulaciones realizadas, las pruebas de bondad de ajuste de Rao-Scott mostraron un mejor desempeño que el resto de las pruebas. La prueba de Rao-Scott con corrección de primer orden (Rao-Scott 1) disminuye el error tipo I con respecto a las pruebas chi-cuadrado de Pearson, Pearson ponderada y Wald. A pesar de esto, la corrección que hace esta prueba no es suficiente, pues aún los valores de $\hat{\alpha}$ son bastantes diferentes del valor nominal ($\alpha = 0,05$).

La prueba de Rao y Scott con corrección de segundo orden (Rao-Scott 2) muestra los valores de $\hat{\alpha}$ más bajos, oscilando éstos entre 0,0583 y 0,1167.

Para observar mejor el comportamiento del error tipo I, se realizaron gráficos con los valores de $\hat{\alpha}$ obtenidos en cada prueba de bondad de

CUADRO 1

Error tipo I estimado de las pruebas de bondad de ajuste.
Estimated Type I error for the goodness of fit tests.

Base de Datos	Variable	k*	np**	$\hat{\alpha}$ de las pruebas de Bondad de Ajuste				
				Pearson	Pearson Ponderada	Wald	Rao-Scott 1	Rao-Scott 2
1	CD: categoría diamétrica	3	5	0,6518	0,2317	0,2019	0,1482	0,1167
			10	0,9001	0,2312	0,1780	0,1275	0,1161
			15	0,9719	0,2348	0,1673	0,1249	0,1147
1	CD: categoría diamétrica	5	5	0,5858	0,2652	0,3711	0,2066	0,1157
			10	0,8585	0,3018	0,3248	0,2084	0,1126
			15	0,9632	0,3140	0,2795	0,1636	0,1140
1	CA: categoría diamétrica	10	5	0,5203	0,3699	0,4590	0,2031	0,1031
			10	0,7883	0,3829	0,5673	0,1847	0,0976
			15	0,9246	0,4154	0,5666	0,1695	0,0897
2	CD: categoría diamétrica	3	5	1	0,1889	0,0922	0,0858	0,0668
			10	1	0,1832	0,0823	0,0805	0,0605
			15	1	0,1824	0,0856	0,0772	0,0583
2	CD: categoría diamétrica	5	5	1	0,1814	0,1654	0,1143	0,0700
			10	1	0,1810	0,1236	0,1122	0,0851
			15	1	0,1747	0,1162	0,1078	0,0617
2	CD: categoría diamétrica	10	5	1	0,1821	0,2491	0,1556	0,0826
			10	1	0,1832	0,1963	0,1503	0,0794
			15	1	0,1829	0,1609	0,1539	0,0810
2	E: Especie	3	5	0,5273	0,2860	0,2410	0,1325	0,1037
			10	0,7129	0,3002	0,2180	0,1381	0,1022
			15	0,8258	0,3064	0,2020	0,1321	0,0994

* k: número de categorías.

** np: número de puntos muestrales.

ajuste y para todos los casos simulados. En la figura 1 se muestra un gráfico representativo, donde se puede observar el error tipo I de las pruebas de bondad de ajuste aplicadas a la variable categoría diamétrica (k=10) de la base de datos 2.

Se puede notar en la figura 1 el comportamiento bastante deficiente de la prueba chi-cuadrado de Pearson, pues el error tipo I estimado es igual a 1 para esta situación particular. Igualmente, es posible observar como el error tipo I disminuye al utilizar las pruebas que toman en cuenta el diseño muestral, siendo la prueba de Rao-Scott

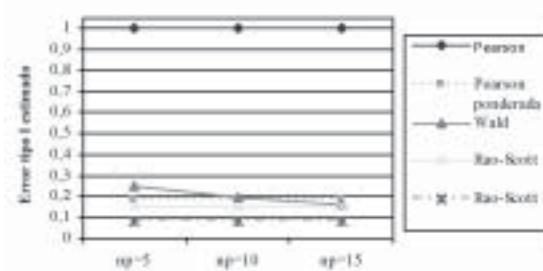


Figura 1. Error tipo I estimado de las pruebas de bondad de ajuste. Variable categoría diamétrica (k=10, Base de Datos 2).

Estimated Type I error for the goodness of fit test. Variable diametric category (k = 10, Database 2).

CUADRO 2

Error tipo I estimado de las pruebas de independencia.
Estimated Type I error for the independence tests.

Base de Datos	Variable	r x c*	np**	$\hat{\alpha}$ de las pruebas de Bondad de Ajuste				
				Pearson	Pearson Ponderada	Wald	Rao-Scott 1	Rao-Scott 2
1	categoría diamétrica vs. variable binomial	3 x 2	5	0,0947	0,0736	0,0806	0,0698	0,0619
			10	0,1021	0,0811	0,0776	0,0705	0,0612
			15	0,1025	0,0750	0,0712	0,0652	0,0595
1	categoría diamétrica vs. variable multinomial	3 x 3	5	0,1009	0,1013	0,1425	0,0978	0,0722
			10	0,1043	0,1017	0,1236	0,0973	0,0694
			15	0,1061	0,1092	0,1201	0,0962	0,0653
2	categoría diamétrica vs. variable binomial	3 x 2	5	0,0958	0,0953	0,0939	0,0927	0,0724
			10	0,1056	0,0923	0,0927	0,0968	0,0629
			15	0,1228	0,1010	0,0824	0,1034	0,0608
2	categoría diamétrica vs. variable multinomial	3 x 3	5	0,0954	0,0992	0,1593	0,1112	0,0697
			10	0,0948	0,0956	0,1182	0,0992	0,0625
			15	0,0963	0,0986	0,1014	0,0949	0,0611

* r: número de filas de la tabla de contingencia.

c: número de columnas de la tabla de contingencia.

** np: número de puntos muestrales.

con corrección de segundo orden la que muestra un mejor desempeño.

En el cuadro 2 se muestran los resultados obtenidos en las pruebas de independencia. Puede observarse en este cuadro, que la prueba de independencia chi-cuadrado de Pearson presenta un error tipo I entre 0,0947 y 0,1061. Estos valores son mucho menores que los registrados por la prueba de bondad de ajuste chi-cuadrado de Pearson, lo que indica que la prueba de independencia de Pearson es menos susceptible a la violación de los supuestos del muestreo con parcelas de tamaño variable, que la prueba de bondad de ajuste de Pearson.

En las figuras 2 y 3 se presenta en forma gráfica el comportamiento del error tipo I de las pruebas de independencia aplicadas a muestras obtenidas de la base de datos 1 y para tablas de contingencia 3x2 y 3x3, respectivamente. El comportamiento de las pruebas es muy similar cuando se trabaja con la base de datos 2.

En la figura 2 puede notarse que para las tablas de contingencia 3x2, la prueba chi-cuadrado

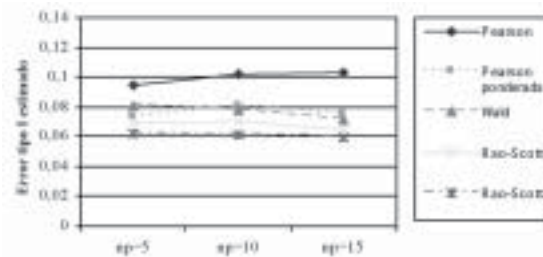


Figura 2. Error tipo I estimado de las pruebas de independencia en tablas de contingencia 3x2. (categoría diamétrica vs. variable binomial).

Estimated Type I error for the Independence test for 2 x 2 contingency tables (diametric category vs. binomial variable).

de Pearson muestra los valores del error tipo I más altos. Se observa además que las correcciones de la prueba de Pearson ponderada por las probabilidades de inclusión y la de Wald disminuyen el error tipo I estimado de la prueba de Pearson, ubicándolo entre 0,07 y 0,09.

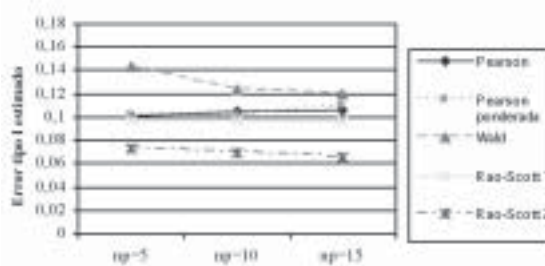


Figura 3. Error tipo I estimado de las pruebas de independencia en tablas de contingencia 3x2 (categoría diamétrica vs. variable multinomial).

Estimated Type I error for the independence test for 3 x 2 contingency tables (diametric category vs. multinomial variable).

En cuanto a las pruebas de Rao-Scott, éstas presentan un error tipo I estimado más cercano al valor nominal ($\alpha=0,05$) que las demás pruebas. Específicamente, la prueba de Rao-Scott con corrección de segundo orden registra los menores valores de $\hat{\alpha}$, estando éstos alrededor de 0,06.

Al analizar el gráfico de la figura 3, que corresponde a tablas de contingencia 3x3, se puede notar que la prueba de Wald es la que presenta el error tipo I estimado más alto. Estos valores son mayores que los obtenidos al usar la prueba de Wald en tablas de contingencia 3x2. Lo anterior hace pensar que la exactitud de esta prueba disminuye al aumentar el tamaño de la tabla, bajo las condiciones experimentales simuladas.

Las pruebas de Pearson, Pearson ponderada y Rao-Scott con corrección de primer orden muestran entre sí un comportamiento similar, registrando valores de $\hat{\alpha}$ muy cercanos a 0,1. Este resultado indica que las correcciones que hacen la prueba de Pearson ponderada y la prueba de Rao-Scott de primer orden no disminuyen el error tipo I de la prueba de independencia chi-cuadrado de Pearson, en tablas de contingencia 3x3. Una vez más, la prueba de Rao-Scott con corrección de segundo orden es la que registra los valores de $\hat{\alpha}$ más cercanos al α nominal (0,05).

Al comparar las figuras 2 y 3 se observa que las pruebas de Pearson, Pearson ponderada y las de Rao-Scott se comportan en forma muy similar en tablas de contingencia 3x2 y 3x3. Por lo tanto, estas pruebas de independencia no muestran una relación directa entre el error tipo I y el tamaño de la tabla de contingencia.

CONCLUSIONES

Los resultados obtenidos en esta investigación indican que la validez de las pruebas de bondad de ajuste e independencia chi-cuadrado de Pearson se ve afectada por la violación de los supuestos de independencia e igual distribución de las observaciones que ocurre en el muestreo con parcelas de tamaño variable.

Las pruebas de bondad de ajuste chi-cuadrado de Pearson, chi-cuadrado de Pearson ponderada por las probabilidades de inclusión, Wald y la de Rao-Scott con corrección de primer orden presentan valores inaceptables para el error tipo I, razón por la cual no se recomienda su uso bajo un muestreo con parcelas de tamaño variable.

En todas las simulaciones realizadas la prueba de bondad de ajuste de Rao-Scott con corrección de segundo orden es la que muestra el mejor desempeño, sin embargo, en algunas situaciones el error tipo I alcanza valores cercanos a 0,12, lo cual no es satisfactorio. Se cree conveniente investigar acerca de otras pruebas que consideren el diseño muestral, para determinar si alguna puede disminuir el error tipo I registrado por la prueba de Rao-Scott con corrección de segundo orden.

La magnitud de la distorsión del error tipo I es mayor en las pruebas de bondad de ajuste que en las pruebas de independencia, lo que indica que estas últimas son menos sensibles a la violación de los supuestos.

En lo que respecta a las pruebas de independencia, no se justifica utilizar la prueba de Pearson ponderada por las probabilidades de inclusión y la de Rao-Scott con corrección de primer orden, ya que presentan valores del error tipo I similares a los obtenidos en la prueba de independencia chi-cuadrado de Pearson. Por lo tanto, no tiene sentido hacer las correcciones pertinentes, si con ello no disminuye el error tipo I.

En cuanto a la prueba de independencia de Wald, ésta presenta el inconveniente de que su error tipo I tiende a aumentar cuando se incrementa el número de celdas de la tabla de contingencia, hecho que indica que esta prueba no es confiable.

La prueba de independencia de Rao-Scott con corrección de segundo orden presenta los valores más bajos del error tipo I, estando muy cerca del valor nominal establecido ($\alpha=0,05$). Es por ello que se recomienda el uso de esta prueba, en situaciones similares a las simuladas en este trabajo.

BIBLIOGRAFIA

- (1) QUINTERO, M. A. Comparación de la distorsión del error tipo I en las pruebas de independencia y de bondad de ajuste, utilizando el muestreo con parcelas de área fija y el método de Bitterlich. Tesis de Maestría. Instituto de Estadística Aplicada y Computación. Facultad de Ciencias Económicas y Sociales. Universidad de Los Andes, Mérida, Venezuela. 2003. 150 p.
- (2) SARNDAL C., B. SWENSSON, J. WRETMAN. *Model Assisted Survey Sampling*. Nueva York: Springer-Verlag. 1992. 694 p.
- (3) FAY, R. A Jackknifed Chi-Squared Test for Complex Samples. *Journal of the American Statistical Association*, 1985, vol. 80, N° 389, p. 148-157.
- (4) STEHMAN, S., S. OVERTON. *Environmental Sampling and Monitoring. Handbook of Statistics*, 1994, vol. 12, p. 263-305.
- (5) HOLT D., A. SCOTT, P. EWINGS. Chi-squared Tests with Survey Data. *Journal of the Royal Statistical Society A*, 1980, vol. 143, p. 303-320.
- (6) SHARON, L. Muestreo: *Diseño y Análisis*. Buenos Aires, Argentina: Internacional Thomson Editores. 2000. 480 p.
- (7) THOMAS D., J. RAO. Small – Simple Comparisons of Level and Power for Simple Goodness-of-Fit Statistics Under Cluster Sampling. *Journal of the American Statistical Association*, 1987, vol. 82, N° 398, p. 630-636.
- (8) RAO J., A. SCOTT. Chi-Squared Tests for Analysis of Categorical Data From Complex Surveys. *Proceedings of The American Statistical Association, Section on Survey Research Methods*, 1979, p. 58-66.
- (9) RAO J., A. SCOTT. The Analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables. *Journal of the American Statistical Association*, 1981, vol. 76, N° 374, p. 221-230.
- (10) RAO J., A. SCOTT. On Chi-squared Tests for Multi-way contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, 1984, vol. 12, p. 46-60.
- (11) ALTHAM, P. Discrete variable analysis for individuals grouped into families. *Biometrika*, 1976, vol. 63, N° 2, p. 263-269.
- (12) COHEN, J. The Distribution of the Chi-Squared Statistic Under Clustered Sampling from Contingency Tables. *Journal of the American Statistical Association*, 1976, vol. 71, N° 355, p. 665-670.
- (13) FELLEGI, I. Aproximate Tests of Independence and Goodness of Fit based on Stratified multistage samples. *Journal of the American Statistical Association*, 1980, vol. 75, p. 261-268.
- (14) NATHAN, G. Test of Independence in Contingency Tables from Stratified Proportional Samples. *Sankhya C*, 1985, vol. 37, p. 77-87.
- (15) LEHTONEN R., E. PAHKINEN. *Practical Method for Design and Analysis of Complex Surveys*. Nueva York: John Wiley & Sons. 1996. 344 p.
- (16) BRICK, J. M., P. BROENE, P. JAMES, J. SEVERYNSE. *A user's Guide to WesVarPc version 2.1*. Maryland, USA: Westat, Inc. 1997. 212 p.
- (17) BITTERLICH, W. *The Relascope Idea, relative Measurements in Forestry*. Londres: Common Wealth Agriculture Bureaux. 1984. 242 p.
- (18) DE VRIES, P. *Sampling Theory for Forest Inventory: A Teach-Yourself Course*. Berlin: Springer-Verlag. 1986. 399 p.
- (19) SCHREUDER, H., T. GREGOIRE, G. WOOD. *Sampling Methods for Multiresource Forest Inventory*. Nueva York: John Wiley and Sons. 1993. 446 p.
- (20) GREGOIRE, T. The Unbiasedness of the Mirage Correction Procedure for Boundary Overlap. *Forest Science*, 1982, vol. 28, N° 3, p. 504-508.

Recibido: 02.03.04

Aceptado: 27.07.04