APLICACIÓN DE DOS PRUEBAS ESTADÍSTICAS DE BONDAD DE AJUSTE EN MUESTRAS COMPLEJAS: UN CASO PRÁCTICO EN EL CAMPO FORESTAL

APPLICATION OF TWO STATISTICAL GOODNESS-OF-FIT TESTS IN COMPLEX SAMPLES: A PRACTICAL CASE IN FOREST FIELD

María A. **Quintero-Méndez** y Mariano J. **Durán-Núñez**

Universidad de Los Andes. Falcultad de Ciencias Forestales y Ambientales. Escuela de Ingeniería Forestal. Mérida, Venezuela (mariaq@ula.ve) (mjdurann.ureach.com)

Resumen

En esta investigación se comparan dos pruebas de bondad de ajuste en términos de su error tipo I: ji-cuadrada de Pearson y Rao-Scott con corrección de segundo orden, aplicadas a datos recolectados mediante técnicas de muestreo que no cumplen los supuestos de independencia e igual probabilidad de inclusión de las observaciones, llamadas muestras complejas. Ambas pruebas se usaron para ajustar categorías diamétricas en una plantación de gmelina (*Gmelina arborea*), aplicando muestreo sistemático con parcelas de área fija y parcelas de área variable (Muestreo de Bitterlich o Parcelas de Radio Variable), mediante técnicas de simulación. La prueba de Rao-Scott con corrección de segundo orden registra un error tipo I más bajo y cercano al valor nominal α que la prueba ji-cuadrada de Pearson, debido a que toma en cuenta los efectos del diseño muestral y corrige la violación de los supuestos. Los resultados obtenidos en esta investigación muestran la inconveniencia de usar la prueba de bondad de ajuste ji-cuadrada de Pearson en datos obtenidos mediante muestreos con parcelas fijas y parcelas de área variable, ampliamente usados en el campo forestal. Por tanto, es necesario usar pruebas estadísticas que consideren la complejidad del diseño muestral, a fin de obtener inferencias válidas.

Palabras clave: *Gmelina arborea*, error tipo I, parcelas de área fija, parcelas de área variable.

Introducción

Muchos de los análisis estadísticos que se aplican a los datos de una muestra requieren que las observaciones sean independientes y que tengan iguales probabilidades de selección (Skinner *et al.*, 1989). Estos supuestos sólo se satisfacen cuando se emplea un muestreo aleatorio simple con reemplazo, y se cumplen aproximadamente en una muestra aleatoria simple sin reemplazo, para una fracción de muestreo pequeña (Sarndal *et al.*, 2003).

Recibido: Julio, 2007. Aprobado: Febrero, 2008.
Publicado como ARTÍCULO en Agrociencia 42: 287-297. 2008.

Abstract

In this research two goodness-of-fit tests are compared in terms of their type I error: Pearson's Chi-square test and Rao-Scott test with correction of second order, applied to data collected using sampling methods that do not fulfill the assumptions of independence and equal probability of inclusion of observations, methods called complex surveys. Both tests were utilized to fit diametric categories in a gmelina plantation (*Gmelina arborea*), applying systematic sampling with fixed area plots and with variable area plots (Bitterlich Sampling or variable radius plot), and employing simulation techniques. The Rao-Scott test with correction of second order registered a lower Type I error, close to the nominal α, when compared to the Pearson Chi-square test, due to the fact that the former takes into account the effects of the sample design and corrects the violation of the assumptions. The results obtained in this research show that the use of Pearson's Chi-square goodness-of-fit test is inappropriate in data obtained applying fixed area and variable area plots, widely used in forestry inventories. Therefore, it is important to use statistical tests that take into account sampling design complexity, in order to achieve valid inferences.

Key words: *Gmelina arborea*, type I error, fixed area plots, variable area plots.

Introduction

Many of the statistical analyses, which are applied to sample data, require the observations to be independent and to have equal probabilities of selection (Skinner *et al.*, 1989). These assumptions are only satisfied when simple random sampling with replacement is employed, and they are fulfilled approximately in a simple random sample without replacement for a small sampling fraction (Sarndal *et al.*, 2003).

In practice, often the used sample designs do not satisfy the assumptions of simple random sampling; some observations can have different selection probabilities, or for logistic reasons, the individuals of a sample form clusters, causing the sample units not to be independent. The whole of observations made using

En la práctica, muchas veces los diseños muestrales usados no satisfacen los supuestos del muestreo aleatorio simple; algunas observaciones pueden tener diferentes probabilidades de selección o, por razones logísticas, los individuos de una muestra forman conglomerados, causando que las unidades muestrales no sean independientes. Al conjunto de observaciones realizadas usando una técnica de muestreo con estas características, se le denomina muestra compleja (Carlson, 1998).

Usualmente los análisis estadísticos de una muestra compleja se hacen como si las observaciones cumplieran con los supuestos del muestreo aleatorio simple (Lee *et al.*, 1989). Es bastante común emplear paquetes estadísticos estándar y no considerar la complejidad del diseño muestral. Pero, ¿se puede ignorar la violación de los supuestos de independencia e igual probabilidad de selección de las observaciones sin afectar la validez de los métodos estadísticos usados?

La prueba ji-cuadrada de Pearson es una de las más usadas para estudiar la bondad de ajuste, por lo que es importante determinar si la violación de los supuestos que ocurre en el muestreo con parcelas de área fija y de área variable, diseños muestrales complejos usados en el campo forestal, afecta su validez estadística. Se ha analizado el comportamiento de la prueba cuando se usa muestreo estratificado y muestreo por conglomerados (Holt *et al.*, 1980; Rao y Scott, 1981; Thomas y Rao, 1987); sin embargo, no se ha estudiado su comportamiento cuando se aplica a datos obtenidos mediante técnicas de muestreo forestal.

Se han propuesto métodos alternativos para probar bondad de ajuste considerando la complejidad del diseño muestral: Fay (1985), Rao y Scott (1979, 1981, 1984), Rai *et al.* (2001). Una prueba de bondad de ajuste con estas características es la de Rao-Scott con corrección de segundo orden; usa un estadístico de prueba que corrige la violación de los supuestos del muestreo aleatorio simple e incorpora el efecto del diseño muestral.

En este trabajo se estudia el comportamiento de las pruebas de bondad de ajuste ji-cuadrada de Pearson y Rao-Scott con corrección de segundo orden en términos del error tipo I, para ajustar categorías diamétricas de gmelina (*Gmelina arborea*) en muestreos con parcelas fijas y parcelas variables. Se intenta demostrar, mediante un ejemplo que utiliza datos reales de una plantación, que una prueba estadística de bondad de ajuste que considere información sobre el diseño muestral permite realizar inferencias más confiables que la prueba clásica ji-cuadrada de Pearson.

the technique of sampling with these characteristics is called complex survey (Carlson, 1998).

Usually, statistical analyses of a complex sample are conducted as if the observations were fulfilling the assumptions of simple random sampling (Lee *et al.*, 1989). It is rather common to employ standard statistical packages and not to take into account the complexity of the sample design. But can the violation of assumptions of independence and probability of equal observation selection be ignored without affecting the validity of the utilized statistical methods?

Pearson's Chi-square test is one of the most used for studying goodness-of-fit; therefore, it is important to determine if the violation of assumptions occurring in the sampling with fixed area and variable area plots, complex sampling designs used in forest field, will affect its statistical validity. It has been analyzed the performance of the test when stratified sampling and cluster sampling are used (Holt *et al.*, 1980; Rao and Scott, 1981; Thomas and Rao, 1987); however, its behavior has not been studied when it is applied to data obtained through forest sampling techniques.

Alternative methods have been proposed for testing goodness-of-fit considering the complexity of sample design: Fay (1985), Rao and Scott (1979, 1981, 1984), Rai *et al.* (2001). One goodness-of-fit test with these characteristics is that of Rao-Scott with correction of second order; it utilizes a test statistic that corrects the violation of the assumptions of simple random sampling and incorporates the effect of the sample design.

In this paper, the performance of Pearson's Chi-square goodness-of-fit test and the Rao-Scott test with correction of second order is studied in terms of type I error, in order to fit diametric categories of gmelina (*Gmelina arborea*) in samplings with fixed and variable plots. It is attempted to prove by an example, based on real data from a plantation, that a statistical goodness-of-fit test, taking into account information on sampling design, allows carrying out more reliable inferences than Pearson's classical Chi-square test.

Materials and Methods

Sampling with fixed area plots

This sampling method is based on the establishment of plots with fixed dimensions and form (rectangular, circular, and square). A tree or any other individual or object one wants to study is included in the sample if it is within the limits of the established plots (Schreuder *et al.*, 1993).

In sampling with fixed area plots, the probability of a tree *u* be included in the sample (*pu*) is given by:

Materiales y Métodos

Muestreo con parcelas de área fija

Este método de muestreo se basa en el establecimiento de parcelas con dimensiones y forma fija (rectangular, circular, o cuadrada). Un árbol o cualquier otro individuo u objeto que se desea estudiar es incluido en la muestra si se encuentra dentro de los límites de las parcelas establecidas (Schreuder *et al.*, 1993).

En el muestreo con parcelas de área fija, la probabilidad de que un árbol u sea incluido en la muestra (pu) está dada por:

$$p_u = a_u / A \tag{1}$$

donde, a_u es el área definida por el conjunto de puntos donde se puede localizar una parcela, tal que el árbol u sea incluido en la muestra; esta área se denomina área de inclusión. A representa el área del bosque, plantación o sitio a muestrear.

Para parcelas muestrales de forma circular es fácil verificar que el área de inclusión de un árbol es un círculo concéntrico a él, con el mismo radio de la parcela muestral (Schreuder *et al.*, 1993). De esta forma $a_u=a$ y es constante para todos los árboles, por ende todos los individuos tienen igual probabilidad de ser incluidos en la muestra, cumpliéndose uno de los supuestos del muestreo aleatorio simple.

Sin embargo, en el muestreo con parcelas de área fija no se cumple el supuesto de independencia de las observaciones, ya que una vez elegido el punto donde se aplica el muestreo, los árboles de cada parcela forman conglomerados y existe una correlación espacial entre ellos.

Muestreo con parcelas de área variable

El muestreo con parcelas de área variable, también llamado muestreo con parcelas de radio variable o método de Bitterlich, es una técnica que permite seleccionar árboles de una parcela con una probabilidad proporcional al área de la sección transversal o área basal o, lo que es igual, proporcional al cuadrado del diámetro del árbol (De Vries, 1986).

Si A representa el área del sitio del cual se extraerá la muestra y α es el ángulo de barrido utilizado en el muestreo, la probabilidad (p_u) de que un árbol u de diámetro du sea muestreado desde un punto localizado aleatoriamente es:

$$p_u = \frac{\pi \times d_u^2}{4 \times A \times \text{seno}^2 (\alpha / 2)} \tag{2}$$

Para una demostración matemática de esta ecuación y obtener detalles de esta técnica de muestreo consultar De Vries (1986) y Schreuder *et al.* (1993).

De acuerdo con la ecuación 2, la probabilidad de seleccionar un árbol es proporcional al cuadrado de su diámetro. De esta manera, en el muestreo con parcelas de área variable no se satisface el supuesto de igual probabilidad de selección para todas las

$$p_u = a_u / A \tag{1}$$

where, a_u is the area defined by the set of points, where a plot can be located, so that the tree u may be included in the sample; this area is called inclusion area. A represents the forest area, plantation, or sampling site.

For sample plots of circular form, it is easy to verify that the inclusion area of a tree is a concentric circle around it, with the same radius as the sample plot (Schreuder *et al.*, 1993). This way $a_u=a$ and it is constant for all the trees, hence all the individuals have equal probability of being included in the sample, one of the assumptions of simple random sampling being fulfilled.

However, in sampling of fixed area plots the assumption of observation independence is not fulfilled, since once the point for the sampling application is chosen, the trees of each plot form clusters and there is spatial correlation among them.

Sampling with variable area plots

Sampling with variable area plots, also called sampling with plots of variable radius or Bitterlich method, is a technique which allows selecting trees from a plot with probability proportional to the basal area of the tree, or which is the same, proportional to the square diameter of the tree (DeVries, 1986).

If A represents the area of the site from which the sample will be extracted and α is the angle gauge used in the sampling, the probability (p_u) that a tree u of diameter du may be sampled from a randomly located point is:

$$p_u = \frac{\pi \times d_u^2}{4 \times A \times \text{sine}^2 (\alpha / 2)} \tag{2}$$

For a mathematical demonstration of this equation and in order to obtain details of this sampling technique, DeVries (1986) and Schreuder *et al.* (1993) should be consulted.

According to equation 2, the probability of selecting a tree is proportional to its square diameter. This way, in sampling with variable area plots, the assumption of equal selection probability is not satisfied for all sample units, since the trees with larger diameter have higher probability of being selected. The distance from the tree to the sampling point influences (the selection) as well.

This sampling technique does not satisfy the assumption of observation independence either, since the trees selected at a sampling point form clusters.

Pearson's Chi-square goodness-of-fit test and Rao-Scott test of second order

In order to apply Pearson's Chi-square goodness-of-fit test, the observations were classified in k categories or classes, and it is assumed that they are independent and distributed identically. The null hypothesis is Ho: $p_i = p_{io}$ for i = 1, 2,...,k; where p_i is the proportion of individuals belonging to category i, p_{io} is the theoretical proportion of category i, and k is the number of categories.

unidades muestrales, ya que los árboles con mayor diámetro tienen una probabilidad más alta de ser seleccionados. También influye la distancia del árbol al punto de muestreo.

Esta técnica de muestreo tampoco satisface el supuesto de independencia de las observaciones, ya que los árboles seleccionados en un punto muestral forman conglomerados.

Pruebas de bondad de ajuste ji-cuadrada de Pearson y Rao-Scott de segundo orden

Para aplicar la prueba de bondad de ajuste ji-cuadrada de Pearson las observaciones son clasificadas en k categorías o clases, y se supone que son independientes e idénticamente distribuidas. La hipótesis nula es Ho: $p_i = p_{io}$ para $i = 1, 2,\ldots$,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.

El estadístico de prueba es:

$$X_p^2 = n\sum_{i=1}^{k}\frac{\left(\hat{p}_i - p_{io}\right)^2}{p_{io}} \tag{3}$$

donde $\hat{p}_i$ se obtiene dividiendo el número de individuos de la categoría i observados entre el total de individuos $\left(\hat{p}_i = n_i / n\right)$.

Si las observaciones son independientes y cada individuo tiene la misma probabilidad de ser seleccionado en la muestra, el estadístico χ_p^2 sigue asintóticamente una distribución ji-cuadrada con $k-1$ grados de libertad.

Cuando se utiliza un diseño muestral complejo, el estadístico χ_p^2 no se distribuye $\chi^2_{(k-1)}$, pero tiene una distribución simétrica, y un múltiplo de χ_p^2 podría seguir aproximadamente una distribución χ^2 (Lohr, 2000).

Rao y Scott (1981) determinaron que el estadístico χ_p^2 se distribuye asintóticamente como una suma ponderada $\delta_1 W_1 + \delta_2 W_2 + \ldots\ldots + \delta_{k-1}W_{k-1}$ de variables aleatorias ji-cuadradas W_j, cada una con un grado de libertad. Los pesos δ_j son los valores propios de la matriz de los efectos del diseño generalizados; esta matriz se define como $\mathbf{D} = \mathbf{P}_o^{-1}\mathbf{V}$, donde $\mathbf{P}_o = diag\left(\mathbf{p}_o\right) - \mathbf{p}_o\mathbf{p}_o'$, $\mathbf{p}_o$ es el vector de proporciones teóricas y $\mathbf{V}/n$ es la matriz de covarianzas del vector de proporciones estimadas $\hat{\mathbf{p}}$. Si se utiliza un muestreo aleatorio simple, los valores propios δ_j de la matriz de los efectos del diseño generalizados son iguales a uno. Así, la suma ponderada $\sum_{j=1}^{k-1}\delta_j W_j$ se reduce a una suma de $k-1$ variables aleatorias ji-cuadradas independientes con un grado de libertad, cuya distribución es $\chi^2_{(k-1)}$. Si el diseño muestral es más complejo, los efectos del diseño generalizados δ_j no son iguales a 1; por tanto, la distribución asintótica de la variable aleatoria $\sum_{j=1}^{k-1}\delta_j W_j$ no es $\chi^2_{(k-1)}$ (Lehtonen y Pahkinen, 2004).

Basados en este hecho, Rao y Scott (1981) proponen dos correcciones al estadístico ji-cuadrado de Pearson. La corrección de primer orden ajusta la esperanza, y la de segundo orden la

The test statistic is:

$$X_p^2 = n\sum_{i=1}^{k}\frac{\left(\hat{p}_i - p_{io}\right)^2}{p_{io}} \tag{3}$$

where, $\hat{p}_i$ is obtained by dividing the number of individuals of category i observed by the total of individuals $\left(\hat{p}_i = n_i / n\right)$.

If the observations are independent and each individual has the same probability of being selected in the sample, the χ_p^2 statistic follows asymptotically a Chi-square distribution with k - 1 degree of freedom.

When a complex sample design is used, the χ_p^2 statistic is not distributed $\chi^2_{(k-1)}$, but it has a symmetric distribution, and a multiple of χ_p^2 could approximately follow an χ^2 distribution (Lohr, 2000).

Rao and Scott (1981) determined that the χ_p^2 statistic is distributed asymptotically as an adjusted sum $\delta_1 W_1 + \delta_2 W_2 + \ldots\ldots + \delta_{k-1}W_{k-1}$ of Chi-square random variables W_j, each with one degree of freedom. The δ_j weights are the eigenvalues of the matrix of the generalized design effects; this matrix is defined as $\mathbf{D} = \mathbf{P}_o^{-1}\mathbf{V}$, where $\mathbf{P}_o = diag\left(\mathbf{p}_o\right) - \mathbf{p}_o\mathbf{p}_o'$, $\mathbf{p}_o$ is the vector of theoretical proportions and $\mathbf{V}$/n is the matrix of co-variances of the vector of estimated proportions $\hat{\mathbf{p}}$. If simple random sampling is utilized, the eigenvalues of δ_j of the matrix of generalized design effects are equal to one. Thus, the weighted sum $\sum_{j=1}^{k-1}\delta_j W_j$ is reduced to a sum of $k-1$ chi-squared independent randomized variables with one degree of freedom whose distribution is $\chi^2_{(k-1)}$. If the sample design is more complex, the generalized effects of design δ_j are not equal to 1: therefore, the asymptotic distribution of the random variable $\sum_{j=1}^{k-1}\delta_j W_j$ is not $\chi^2_{(k-1)}$ (Lehtonen and Pahkinen, 2004).

Based on this fact, Rao and Scott (1981) propose two corrections for Pearson's Chi-square statistic. The correction of first order fits expectation, and that of second order fits expectation and asymptotic variance. In this study, the correction of second order is employed, also called Sattlerthwaite correction. The Rao-Scott test with correction of second order uses the statistic:

$$\chi^2_{RS2} = \frac{\dfrac{\chi_p^2}{\hat{\delta}}}{1 + \hat{a}^2} \tag{4}$$

where, χ_p^2 is Pearson's Chi-square statistic; $\hat{\delta}$ is an estimator of the mean $\bar{\delta}$ of eigenvalues δ_j of the matrix of generalized design effects; â is an estimator of the coefficient of δ_j eigenvalue variation.

If the sample design has unequal inclusion probabilities, the χ_p^2 statistic is calculated utilizing estimators of $\hat{p}_i$ proportions weighted by inclusion probabilities:

esperanza y la varianza asintótica. En este estudio se emplea la corrección de segundo orden, también llamada corrección de Satterthwaite. La prueba de Rao-Scott con corrección de segundo orden usa el estadístico:

$$\chi^2_{RS2} = \frac{\frac{\chi^2_p}{\hat{\delta}}}{1+\hat{a}^2} \tag{4}$$

donde, χ^2_p es el estadístico ji-cuadrado de Pearson; $\hat{\delta}$ es un estimador de la media $\bar{\delta}$ de los valores propios δ_j de la matriz de los efectos del diseño generalizados; â es un estimador del coeficiente de variación de los valores propios δ_j.

Si el diseño muestral tiene probabilidades de inclusión desiguales, se calcula el estadístico χ^2_p utilizando estimadores de proporciones $\hat{p}_i$ ponderados por las probabilidades de inclusión:

$$\hat{p}_i = \frac{\sum_{j=1}^{n} W_j Y_{ij}}{\sum_{j=1}^{n} W_j} \quad \text{con} \quad W_j = 1 / \Pi_j \tag{5}$$

donde, Π_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 con valor 1 si el individuo j pertenece a la categoría i y 0 en otro caso.

Las ecuaciones para calcular $\bar{\delta}$ y â^2 son:

$$\hat{\delta} = \frac{n}{(k-1)} \sum_{i=1}^{k} \frac{\hat{v}(\hat{p}_i)}{p_{oi}} \tag{6}$$

siendo $\hat{v}(\hat{p}_i)$ la varianza del estimador de la proporción de la categoría i,

$$\hat{a}^2 = \frac{\sum_{j=1}^{k-1} \hat{\delta}_j^2}{(k-1)\hat{\delta}^2} - 1 \tag{7}$$

$\sum_{j=1}^{k-1} \hat{\delta}_j^2$ se calcula de la siguiente manera:

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

donde, $\hat{V}^2(\hat{p}_i, \hat{p}_j)$ es 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. Los valores de $\hat{V}^2(\hat{p}_i, \hat{p}_j)$ se pueden estimar utilizando una técnica de re-muestreo como la técnica de replicación balanceada, bootstrap o Jackknife. En esta investigación se utilizó la técnica de Jackknife.

$$\hat{p}_i = \frac{\sum_{j=1}^{n} W_j Y_{ij}}{\sum_{j=1}^{n} W_j} \quad \text{with} \quad W_j = 1 / \Pi_j \tag{5}$$

where, Π_j is the probability that individual j may be selected in the sample (inclusion probability of individual j), and Y_{ij} is an indicator variable with value 1, if individual j belongs to category i and 0 in another case.

The equations to calculate $\bar{\delta}$ and â^2 are:

$$\hat{\delta} = \frac{n}{(k-1)} \sum_{i=1}^{k} \frac{\hat{v}(\hat{p}_i)}{p_{oi}} \tag{6}$$

$\hat{v}(\hat{p}_i)$ being the variance of the proportion estimator of category i,

$$\hat{a}^2 = \frac{\sum_{j=1}^{k-1} \hat{\delta}_j^2}{(k-1)\hat{\delta}^2} - 1 \tag{7}$$

$\sum_{j=1}^{k-1} \hat{\delta}_j^2$ is calculated as follows:

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

where, $\hat{V}^2(\hat{p}_i, \hat{p}_j)$ is the element i, j of the covariance matrix of the proportion estimators; p_{oi} and p_{oj} correspond to the theoretical proportions for categories i and j. The values $\hat{V}^2(\hat{p}_i, \hat{p}_j)$ can be estimated using a re-sampling technique like the technique of balanced replication, bootstrap, or Jackknife. In this research the Jackknife technique was used.

χ^2_{RS2} statistic is distributed approximately as chi-square with $\frac{K-1}{1+\hat{a}^2}$ degrees of freedom. In most of the cases the degrees of freedom are a decimal number; therefore, they are rounded off to the next inferior integer. The theoretical foundations of this test may be found in Rao and Scott (1979, 1981, 1984), Sarndal *et al.* (2003), and Lehtonen and Pahkinen (2004).

Data collection

In this paper, data of a gmelina plantation located in the farm El Hierro, State of Portuguesa, property of the company SMURFIT, Cartón de Venezuela, were used. In the plantation, census was taken of a 4.8 ha area with 4841 trees. Each tree was located in coordinate axes (x,y), and its diameter at breast height (DBH=DAP) was measured. Three categorical diameter variables were established, utilizing 5, 10, and 15 categories, used in the goodness-of-fit tests. Thus, for example, in order to establish a variable with 5 diametric

El estadístico χ^2_{RS2} se distribuye aproximadamente como una ji-cuadrada con $\frac{K-1}{1+\hat{a}^2}$ grados de libertad. En la mayoría de los casos los grados de libertad son un número decimal, por lo que se redondean al entero inferior más próximo. Los fundamentos teóricos de esta prueba pueden verse en Rao y Scott (1979, 1981, 1984), Sarndal *et al*. (2003) y Lehtonen y Pahkinen (2004).

Recolección de datos

En este trabajo se utilizaron datos de una plantación de gmelina ubicada en la finca El Hierro, Estado Portuguesa, propiedad de la empresa SMURFIT Cartón de Venezuela. En la plantación se hizo un censo de un área de 4.8 ha, con 4841 árboles. Cada árbol se ubicó en un eje cartesiano (x,y), y se midió el diámetro a la altura de pecho (DAP). Se establecieron tres variables categóricas del diámetro usando 5, 10 y 15 categorías, usadas en las pruebas de bondad de ajuste. Así, por ejemplo, para establecer una variable con 5 categorías diamétricas se dividió el rango de valores de DAP en 5 intervalos, y a cada árbol se le asignó un número del 1 al 5 dependiendo del intervalo donde se ubica su DAP. De manera similar se procedió en el caso de 10 y 15 categorías.

Obtención de muestras y aplicación de las pruebas de bondad de ajuste mediante simulación

Se construyeron dos programas de computación que permiten simular la obtención de muestras usando parcelas de área fija y parcelas de área variable. Con los datos de una muestra, estos programas aplican a la variable categoría diamétrica las pruebas de bondad de ajuste ji-cuadrada de Pearson y Rao-Scott con corrección de segundo orden. Para implementar los programas se utilizó el lenguaje GAUSS, versión 3.1.4.

Los programas de simulación requieren como entrada 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 la categoría diamétrica. También es necesario indicar las dimensiones del área a muestrear, el número de puntos de muestreo, el radio de las parcelas muestrales (para el muestreo con parcelas de área fija) y el ángulo de barrido (para el muestreo con parcelas de área variable).

Para simular el muestreo con parcelas de área fija se trabajó con parcelas circulares de radio igual a 10 m, y los puntos iniciales de muestreo se seleccionaron con un diseño sistemático. Se hicieron pruebas con 5 y 8 puntos, correspondientes a una intensidad de muestreo de 3 y 5%. En la simulación del muestreo de parcelas de área variable se tomaron los mismos puntos iniciales utilizados en la simulación del muestreo con parcelas de área fija, para posteriormente realizar comparaciones entre los dos tipos de muestreo.

Los programas aplican el método Mirage para corregir el llamado efecto de borde en los límites del área a muestrear, y el sesgo que producen. El efecto de borde supone que un árbol ubicado cerca de los límites de la parcela puede tener una probabilidad de

categories, the range of DAP values was divided into 5 intervals, and to each tree, a number from 1 to 5 was assigned depending on the interval where its DAP is situated. A similar procedure was followed in the case of 10 and 15 categories.

Samples obtention and application of goodness-of-fit tests by simulation

Two computer programs were constructed which allow simulating to obtain samples using fixed area and variable area plots. With the data of a sample these programs apply to the diametric category variable the chi-square goodness-of-fit tests of Pearson and Rao Scott with correction of second order. GAUSS language, version 3.1.4, was used to implement the programs.

As an entrance, the simulation programs require a data base with information about the tree population of the area where the sample will be extracted, which must include the location of each tree expressed as coordinate (x,y), DAP, and diametric category. It is also necessary to indicate the dimensions of the area from which samples are to be taken, the number of sampling points, the radius of sample plots (for sampling with fixed area plots), and the angle gauge (for sampling with variable area plots).

To simulate sampling with fixed area plots, circular plots with radius of 10 m were employed, and initial sampling points were selected with systematic design. Tests were made with 5 and 8 points corresponding to a sampling intensity of 3 and 5 %. At the simulation of sampling with variable area plots, the same initial points used for sampling simulation with fixed area plots were taken in order to subsequently carry out comparisons between the two sampling types.

The programs apply the Mirage method to correct the so-called border effect in the limits of the sampling area, and the skewness they produce. The border effect assumes that a tree situated near the limits of the plot may have a probability of being included in the sample different from a similar tree, located in the center of the plot (Schreuder *et al*., 1993).

Measuring of the performance of goodness-of-fit tests

In order to evaluate the performance of the two considered goodness-of-fit tests, the Type I error is estimated which, if differs significantly from the established theoretical level, it is considered that the test is not robust and therefore, the validity of the obtained results cannot be guaranteed. One, of the most used criteria of robustness is that of Bradley (1978), which establishes that a test is robust when the empirical rates of type I error are in the interval $[0.5\alpha, 1.5\alpha]$; thus, for $\alpha = 0.05$ this criterion of robustness requires that the estimated type I error must be between 0.025 and 0.075, otherwise, the test cannot be considered robust.

Estimate of type I error

The type I error was estimated by simulation techniques or the Monte Carlo method. For this, the program was executed r times, ser incluido en la muestra diferente a la de un árbol similar situado en el centro de la parcela (Schreuder *et al.*, 1993).

Medición del desempeño de la pruebas de bondad de ajuste

Para medir el desempeño de las dos pruebas de bondad de ajuste consideradas, se estima el error tipo I el cual, si difiere significativamente del nivel nominal (teórico) establecido, se considera que la prueba no es robusta y, por tanto, no puede garantizarse la validez de los resultados obtenidos. Uno de los criterios de robustez más usados es el de Bradley (1978), el cual establece que una prueba es robusta si las tasas empíricas de error tipo I están en el intervalo $[0.5\alpha, 1.5\alpha]$; así, para α=0.05 este criterio de robustez requiere que el error tipo I estimado esté entre 0.025 y 0.075, en otro caso, la prueba no puede considerarse robusta.

Estimación del error tipo I

El error tipo I se estimó mediante técnicas de simulación o método Monte Carlo. Para ello, el programa se ejecutó r veces, lo que implica que se obtienen r muestras y que a cada una de ellas se le aplican las pruebas para determinar si las proporciones de cada categoría diamétrica en la muestra (p_i) son iguales a las proporciones de cada categoría diamétrica en la población (p_{io}), entendiéndose por tal el conjunto de todos los árboles en la plantación.

Al finalizar las r corridas, el programa estima el error tipo I de las dos pruebas estadísticas, con la ecuación:

$$\hat{\alpha} = \frac{X}{r} \tag{9}$$

donde, $\hat{\alpha}$ es el error tipo I estimado; X el número de veces que se rechazó Ho siendo verdadera; r es el número de replicaciones. En todas las simulaciones se utilizaron 5000 replicaciones.

Una vez calculado el valor de $\hat{\alpha}$ para las pruebas de bondad de ajuste, se compara con el nominal establecido (0.05).

Comparación de las pruebas de bondad de ajuste

Se usó la Prueba de Mc Nemar (Conover, 1980) para comparar el error tipo I de las pruebas y establecer si hay diferencias significativas entre ambas.

Resultados y Discusión

Los valores del error tipo I estimado ($\hat{\alpha}$) de las pruebas aplicadas en categorías diamétricas de gmelina, se muestran en los Cuadros 1 y 2. Se presentan los resultados obtenidos para los dos tipos de muestreo utilizados y diferentes condiciones experimentales, así como los valores del estadístico de Mc Nemar obtenidos al comparar las dos pruebas de bondad de ajuste.

which implies that r samples are obtained and that to each of them, tests are applied to determine if the proportions of each diametric category in the sample (p_i) are equal to those of each diametric category in the population (p_{io}), understanding by this, all the trees in the plantation.

Finishing the r runs, the program estimates the type I error of the two statistical tests with the equation:

$$\hat{\alpha} = \frac{X}{r} \tag{9}$$

where, $\hat{\alpha}$ is the estimated type I error; X the number of times Ho was rejected being true; r is the number of replications. In all the simulations 5000 replications were used.

Once the value of $\hat{\alpha}$ for goodness-of-fit tests calculated, it is compared with the established nominal alpha (0.05).

Comparison of goodness-of-fit tests

The McNemar Test (Conover, 1980) was used to compare type I error of the tests and to establish if there are significant differences between both.

Results and Discussion

The values of type I error estimated ($\hat{\alpha}$) of the tests applied in diametric categories of gmelina are shown in Tables 1 and 2. The results obtained for the two sampling types utilized and different experimental conditions are shown, as well as the statistical values of McNemar, obtained at comparing the two goodness-of-fit tests.

In all the simulated situations Pearson's Chi-square test has a higher type I error than that of Rao-Scott with correction of second order. The values of Mc Nemar's statistic are lower than −2.33, indicating highly significant differences between the results of Pearson's Chi-square goodness-of-fit test and those of Rao-Scott's test with correction of second order.

In Table 1 it is shown that when fixed area plots are used, the Chi-square test of Pearson registers values of $\hat{\alpha}$ fluctuating between 0.0836 and 0.1470, moving away from the nominal value (α=0.05). In sampling with variable area plots (Table 2), Pearson's chi-square test presents very high values of $\hat{\alpha}$ (0.7750 to 0.9920), a type I error unacceptable for a test of hypothesis. These values, so high for an estimated type I error, are basically due to the proper characteristics of sampling with variable area plots, since this method favors the selection of trees with larger diameters at breast height, to estimate the total base area and the total volume being among its objectives; and the trees with the highest diametric categories are the ones that render the greatest contribution to the estimate of

Cuadro 1. Error tipo I estimado para las pruebas de bondad de ajuste para α=0.05 aplicadas a categorías diamétricas de gmelina, usando parcelas de área fija.
Table 1. Estimated type I error for goodness-of-fit tests for α=0.05 applied to diametric categories of gmelina, using fixed area plots.

Número de categorías diamétricas (k)	Número de puntos de muestreo (n)	Error tipo I estimado ($\hat{\alpha}$)		Estadístico de Mc Nemar
		χ^2 Pearson	Rao-Scott 2° orden	
5	5	0.0910	0.0760	−4.7†
5	8	0.0836	0.0642	−5.5†
10	5	0.0940	0.0716	−5.9†
10	8	0.0854	0.0606	−6.5†
15	5	0.1470	0.0834	−5.6†
15	8	0.1092	0.0712	−15.0†

† Diferencias altamente significativas ($p \leq 0.01$)

Cuadro 2. Error tipo I estimado para las pruebas de bondad de ajuste aplicadas a categorías diamétricas de gmelina con α=0.05, usando parcelas de área variable.
Table 2. Estimated type I error for goodness-of-fit tests applied to diametric categories of gmelina with α=0.05, using variable area plots.

Número de categorías diamétricas (k)	Número de puntos de muestreo (n)	Error tipo I estimado ($\hat{\alpha}$)		Estadístico de Mc Nemar
		χ^2 Pearson	Rao-Scott 2° orden	
5	5	0.9920	0.0896	−67.1†
5	8	0.9302	0.0746	−65.3†
10	5	0.9720	0.0708	−67.1†
10	8	0.8650	0.0552	−63.7†
15	5	0.9528	0.0610	−66.7†
15	8	0.7750	0.0576	−59.8†

† Diferencias altamente significativas ($p \leq 0.01$).

En todas las situaciones simuladas la prueba ji-cuadrada de Pearson tiene un error tipo I más alto que la de Rao-Scott con corrección de segundo orden. Los valores del estadístico de McNemar son menores a −2.33, indicando diferencias altamente significativas entre los resultados de la prueba de bondad de ajuste ji-cuadrada de Pearson y los de la prueba de Rao-Scott con corrección de segundo orden.

En el Cuadro 1 se muestra que cuando se usan parcelas de área fija la prueba ji-cuadrada de Pearson registra valores de $\hat{\alpha}$ que oscilan entre 0.0836 y 0.1470, alejándose del valor nominal (α=0.05). En el muestreo con parcelas de área variable (Cuadro 2), la prueba ji-cuadrada de Pearson presenta valores muy altos de $\hat{\alpha}$ (0.7750 a 0.9920), un error tipo I inaceptable para una prueba de hipótesis. Estos valores tan altos para el error tipo I estimado se deben fundamentalmente a las características propias del muestreo con parcelas de área variable, ya que este método favorece la selección de árboles con diámetros mayores a la altura de pecho, y uno de sus objetivos es estimar el área basal total y el volumen total, donde los árboles de las categorías diamétricas superiores hacen el mayor aporte para estimar estos parámetros. Así, al analizar los resultados de las simulaciones se observó

these parameters. Thus, analyzing the results of the simulations, it was observed that the lower diametric categories were not represented in the samples, or the number of trees included by these categories was small, causing that the proportions of the diametric categories in a sample (p_i) were very different from the proportions registered in population (p_{io}). Therefore, in most cases the rejection of a true null hypothesis is produced, and type I error was very high. The difference to sampling of fixed area plots basically lies here, where all the trees located within the plot perimeter are included, so that the trees of a certain diametric class are sampled proportionally to the frequency in the forest of this particular class of trees.

Pearson's Chi-square test has a distortion of the estimated type I error with respect to the nominal value (α=0.05) much higher in the sampling with variable area plots than in that of fixed area plots. Besides, in the sampling with fixed area plots, only the assumptions of independence of observations is violated, whereas in that of variable area plots, the assumption of independence and equal distribution of observations are violated.

In Tables 1 and 2 the Chi-square tests of Pearson and Rao-Scott with correction of second order show que las categorías diamétricas inferiores no estaban representadas en las muestras, o el número de árboles incluidos de esas categorías era pequeño, ocasionando que las proporciones de las categorías diamétricas en una muestra (p_i) fueran muy diferentes de las proporciones registradas en la población (p_{io}). Por tanto, en la mayoría de los casos se rechazó una hipótesis nula verdadera y el error tipo I estimado fue muy alto. En ésto radica básicamente la diferencia con el muestreo de parcelas de área fija, donde se incluyen todos los árboles ubicados dentro del perímetro de la parcela, de modo que los árboles de cierta clase diamétrica se muestrean en forma proporcional a la frecuencia en el bosque de esa clase particular de árboles.

La prueba ji-cuadrada de Pearson presenta una distorsión del error tipo I estimado con respecto al valor nominal (α=0.05) mucho mayor en el muestreo con parcelas de área variable que en el de parcelas de área fija. Además, en el muestreo con parcelas de área fija se viola sólo el supuesto de independencia de las observaciones, mientras que en el de parcelas de área variable se violan los supuestos de independencia e igual distribución de las observaciones.

En los Cuadros 1 y 2 las pruebas ji-cuadrada de Pearson y Rao-Scott con corrección de segundo orden muestran variaciones en su desempeño en relación con el tamaño de muestra. En ambas pruebas el error tipo I estimado disminuye cuando aumenta la cantidad de puntos muestrales, para un número fijo de categorías diamétricas. Éste es el comportamiento esperado, puesto que en muestras de tamaño mayor las categorías diamétricas están mejor representadas, por lo que se rechaza en menor proporción la hipótesis nula.

Si se mantiene constante el número de puntos muestrales, se observa que en el muestreo con parcelas de área fija (Cuadro 1) el valor de $\hat{\alpha}$ crece cuando aumenta el número de categorías diamétricas. En el muestreo con parcelas de tamaño variable (Cuadro 2), el error tipo I estimado es menor para un número mayor de categorías.

La prueba de Rao-Scott con corrección de segundo orden, en parcelas fijas y en parcelas variables, presenta valores del error tipo I estimado menores y más cercanos a 0.05 que la prueba ji-cuadrada de Pearson. Los resultados de la prueba de Rao-Scott con las correcciones, basadas en la esperanza y la varianza asintótica además de la inclusión de los efectos del diseño muestral, disminuyen considerablemente el error tipo I.

Rao y Scott (1981) también encontraron diferencias importantes entre las dos pruebas de bondad de ajuste al estudiar el efecto de la estratificación y de los conglomerados sobre la distribución asintótica del

variations in their performance related to sample size. In both tests, the estimated type I error diminishes when the amount of sample points increases for a fixed number of diametric categories. This is the expected behavior, given that in samples of larger size the diametric categories are better represented, which causes that the null hypothesis is rejected to a lesser proportion.

If the number of sampling points is kept constant, it is observed that in sampling with fixed area plots (Table 1), value $\hat{\alpha}$ grows when the number of diametric categories increases. In sampling with variable size plots (Table 2) the estimated type I error is lower for a larger number of categories.

Rao-Scott's test with correction of second order, in fixed area as well as in variable area plots, presents lower values of the estimated type I error, and closer to 0.05 than Pearson's chi-square test. The results of the Rao-Scott test with the corrections, based on expectation and asymptotic variance, as well as the inclusion of the sample design effect, diminish the type I error considerably.

Rao and Scott (1981) also found important differences between the two goodness-of-fit tests, when studying the effect of stratification and clusters on asymptotic distribution of Pearson's Chi-square statistic. Likewise, Thomas and Rao (1987) reported high significance levels for Pearson's Chi-square test with sampling by clusters, and rather good results for the Rao-Scott test. Nevertheless, in none of these studies, Pearson's Chi-square goodness-of-fit test showed such a high type I error as when sampling by variable area plots. The performance of Rao Scott's test in terms of type I error, reported in the mentioned works, is better than the one recorded in the simulations carried out in this study, since in most of the experiments the authors obtained significance levels equal to the theoretical value, while in this study the estimated type I error was always higher than the nominal or theoretical level. However, only in 3 of the simulated cases (20%) the type I error of the Rao-Scott test surpasses the superior limit of Bradley's interval. Therefore, it can be stated that in most situations this goodness-of-fit test proved to be robust. The differences between the values of $\hat{\alpha}$ obtained for the Rao-Scott test with correction of second order, and those reported in other studies, are consequences of the differences between the experimental conditions, the characteristics of the population, and the sampling designs applied.

CONCLUSIONS

The results obtained in this research indicate that the validity of Pearson's chi-squared goodness-of-fit estadístico ji-cuadrado de Pearson. Igualmente, Thomas y Rao (1987) reportaron niveles de significancia altos para la prueba ji-cuadrada de Pearson con el muestreo por conglomerados, y resultados bastante buenos para la prueba de Rao-Scott. Sin embargo, en ninguno de estos trabajos la prueba de bondad de ajuste ji-cuadrada de Pearson mostró un error tipo I tan alto como al muestrear por parcelas variable. El desempeño de la prueba de Rao-Scott en términos del error tipo I reportado en los trabajos mencionados es mejor que el registrado en las simulaciones realizadas en este trabajo, ya que en la mayoría de los experimentos los autores obtuvieron niveles de significancia iguales al valor teórico, mientras que en el presente estudio el error tipo I estimado siempre fue mayor al nivel nominal o teórico. No obstante, sólo en 3 (20%) de los casos simulados el error tipo I de la prueba de Rao-Scott supera el límite superior del intervalo de Bradley. Por tanto, puede decirse que en la mayoría de las situaciones esta prueba de bondad de ajuste se mostró robusta. Las discrepancias entre los valores de $\hat{\alpha}$ obtenidos para la prueba de Rao-Scott con corrección de segundo orden y los reportados en otras investigaciones son consecuencia de las diferencias entre las condiciones experimentales, las características de la población y los diseños muestrales aplicados.

Conclusiones

Los resultados obtenidos en esta investigación indican que la validez de la prueba de bondad de ajuste ji-cuadrada de Pearson es influenciada por la violación de los supuestos de independencia e igual distribución de las observaciones que ocurre en el muestreo con parcelas de área variable. Lo mismo sucede cuando se utiliza el muestreo con parcelas de área fija. La violación del supuesto de independencia hace que la prueba ji-cuadrada de Pearson tenga un error tipo I diferente del valor nominal establecido. La magnitud de la distorsión del error tipo I de estas pruebas varía de un muestreo a otro, siendo mayor cuando se usan parcelas de tamaño variable.

Los niveles de significación altos para la prueba de bondad de ajuste ji-cuadrada de Pearson cuando se aplican en muestras obtenidas con parcelas fijas y parcelas variables puede conducir a conclusiones erróneas acerca de la población en estudio, razón por la cual no se recomienda su uso. Es conveniente aplicar una prueba de bondad de ajuste que corrija las violaciones de los supuestos de independencia e igual probabilidad de inclusión para todos los individuos. Una de estas pruebas es la de Rao-Scott con corrección de segundo orden, que en esta investigación registró valores

test is affected by the violation of the assumptions of independence and equal distribution of observations, which occurs in sampling with variable area plots. The same happens when sampling with fixed area plots is used. Violation of independence assumption causes Pearson's Chi-square test to have a type I error, different from the established nominal value. The magnitude of the distortion of the type I error of these tests varies from one sampling to another, the greatest occurring when plots of variable size are used.

The high significance levels for Pearson's chi-square goodness-of-fit test, when applied in samples obtained with fixed and variable plots, may lead to erroneous conclusions about the population under study, reason why their use is not recommended. It is convenient to apply a goodness-of-fit test that may correct the violations of assumptions of independence and equal inclusion probability for all the individuals. One of these tests is that of Rao-Scott with second order correction, which in this research registered values of the type I error lower than those of Pearson's chi-square test and in many cases rather close to the nominal value.

The results of this study show the importance of considering the complexity of sample design, used when data of a complex sample are analyzed.

—End of the English version—

del error tipo I menores que la prueba ji-cuadrada de Pearson y en muchos casos bastante cercanos al valor nominal.

Los resultados de este trabajo muestran la importancia de considerar la complejidad del diseño muestral utilizado cuando se analizan datos de una muestra compleja.

Literatura Citada

Bradley, J. V. 1978. Robustness? The Br. J. Mathematical & Statistical Psychology 31: 144-152.

Carlson, B. L. 1998. Software for statistical analysis of sample survey data. *In*: Armitage P., and T. Colton (eds). Encyclopedia of Biostatistics. John Wiley and Sons. New York. pp: 4160-4167.

Conover, W. J. 1980. Practical Nonparametric Statistics. 2nd. ed. John Wiley & Son. New York. 493 p.

De Vries, P. 1986. Sampling Theory for Forest Inventory. Springer-Verlag. Nueva York. 399 p.

Fay, R. E. 1985. A jacknifed Chi-squared test for complex samples. JASA 80: 148-157.

Fellegi, I. 1980. Aproximate tests of independence and goodness of fit based on stratified multistage samples. JASA. 75: 261-268.

Holt, D., A. J. Scott, and P. D. Ewings. 1980. Chi-squared tests with survey data. JRSS, A. 143: 303-320.

Lee, E. S., R. N. Forthofer, and R. J. Lorimer. 1989. Analyzing Complex Survey Data. Sage University Paper series on Quantitative Applications in the Social Sciences 07-071. Sage Publications. Beverly Hills, USA. 79 p.

Lehtonen, R., and E. Pahkinen. 2004. Practical Method for Design and Analysis of Complex Surveys. Second Edition. John Wiley & Sons. New York. 320 p.

Lohr, S. 2000. Muestreo: Diseño y Análisis. Internacional Thomson Editores. Buenos Aires, Argentina. 420 p.

Nathan, G. 1975. Test of independence in contingency tables from stratified proporcional samples. Sankhya C 37: 77-87.

Rai, A., A. K. Srivastava, and H.C. Gupta. 2001. Small sample comparison of modified chi-square test statistics for survey data. Biometrical J. 43: 483-495.

Rao, J. N .K, and A. J. Scott. 1979. Chi-Squared tests for analysis of categorical data from complex surveys. Proceedings of The American Statistical Association, Section on Survey Research Methods.pp: 58-66.

Rao, J. N .K, and A. J. Scott. 1981. The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. JASA 76: 221-230.

Rao, J. N .K, and A. J. Scott. 1984. On Chi-squared tests for Multi-way contingency tables with cell proportions estimated from survey data. The Annals of Statistics 12: 46-60.

Sarndal, C. E., B. Swensson, and J. Wretman. 2003. Model Assisted Survey Sampling. Springer- Verlag. New York. 694 p.

Schreuder, H., T. Gregoire, and G. Wood. 1993. Sampling Methods for Multiresource Forest Inventory. John Wiley and Sons. New York. 446 p.

Skinner, C. J., P. Holt, and T. M. K. Smith. 1989. Analysis of Complex Surveys. John Wiley & Sons. Chicester. 328 p.

Thomas, D. R., and J. N. K. Rao.1987. Small–simple comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. JASA 82: 630-636.