

# Comparación de intervalos de confianza para la distribución multinomial

Difariney González Gómez

Director: Juan Carlos Correa Morales  
Ph.D University of Kentucky  
Profesor Asociado, Escuela de Estadística  
Universidad Nacional de Colombia

Trabajo presentado como requisito para optar  
al título de Magister en Estadística

Escuela de Estadística  
Facultad de Ciencias  
Universidad Nacional de Colombia  
Sede Medellín  
2010

*A mi mamá que desde el cielo me cuida, a mi papá y sobrino que son mi apoyo incondicional, a mi director de tesis por su paciencia y apoyo, a todos mil gracias por confiar en mí.*

## Resumen

Uno de los problemas más importantes en la inferencia estadística es encontrar los intervalos de confianza para los parámetros de la distribución multinomial. En este trabajo se evalúan y se comparan diferentes metodologías propuestas en la literatura, analizando para cada intervalo el nivel de confianza, la longitud y el índice, el cual es una combinación de los dos conceptos anteriores. La comparación de los intervalos se desarrolla mediante simulaciones utilizando el paquete estadístico R. En este proceso, además de la verificación de conclusiones conocidas se determinan aspectos relevantes como es el caso de los intervalos encontrados con el método del Teorema del Límite Central, ya que en algunos libros de estadística básica, los proponen como la metodología más utilizada y de acuerdo a este análisis no presentan el mejor desempeño.

Palabras clave: Distribución Multinomial, Intervalos de Confianza, Estimación.

## Abstract

One of the most important problems in statistical inference is to find confidence intervals for parameters of multinomial distribution. In this paper was evaluated and compared different methodologies proposed in the literature, analyzing for each interval the confidence level, length and index, which is a combination of the two previous concepts. Comparison of the intervals is developed through simulations using the statistical package R. In this process, besides the verification of conclusions known determining relevant aspects such as the intervals found with the method of the Central Limit Theorem as in some books of the basic statistical methodology proposed as the most used and according to this analysis does not present the best performance.

Keywords: Multinomial Distribution, Confidence intervals, Estimation.

# Índice general

<b>1. Introducción</b>	<b>8</b>
1.1. Planteamiento del problema . . . . .	9
<b>2. Marco Teórico</b>	<b>10</b>
2.1. Modelo multinomial . . . . .	10
2.2. Intervalos de confianza . . . . .	11
2.2.1. Estimación del intervalo bayesiano . . . . .	11
2.2.2. Intervalos basados en el método de Sison y Glaz . . . . .	12
2.2.3. Intervalo de Quesenberry y Hurst . . . . .	14
2.2.4. Método basado en el Teorema del Límite Central . . . . .	14
2.2.5. Intervalos basados en la Razón de Verosimilitud Relativa . . . . .	15
2.2.6. Método exacto basado en la $F$ . . . . .	15
2.2.7. Intervalo de Goodman . . . . .	16
2.2.8. Intervalo de Bailey . . . . .	16
2.2.9. Intervalo de Fitzpatrick y Scott . . . . .	17
2.3. Entropía como medida de polarización en la multinomial . . . . .	18
2.4. Estado del Arte . . . . .	18
<b>3. Estudio de Simulación</b>	<b>21</b>
3.1. Metodología . . . . .	21
3.2. Resultados y Conclusiones del estudio de simulación . . . . .	38
<b>4. Aplicación: Base de datos de muertes ocurridas en la ciudad de Medellín en 1996</b>	<b>40</b>
<b>5. Conclusiones y Recomendaciones</b>	<b>43</b>
<b>A. Programas en R</b>	<b>47</b>

# Índice de figuras

3.1. Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 3 . . . . .	32
3.2. Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 3 . . . . .	33
3.3. Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 4 . . . . .	34
3.4. Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 4 y 5 . . . . .	35
3.5. Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 5 y 10 . . . . .	36
3.6. Cluster con índice . . . . .	37
4.1. Intervalos basados en el método exacto de la F . . . . .	42

# Índice de tablas

3.1.	Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 3 . . . . .	23
3.2.	<i>(Continuación)</i> Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 3 . . . . .	24
3.3.	<i>(Continuación)</i> Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 3 . . . . .	25
3.4.	<i>(Continuación)</i> Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 4 . . . . .	26
3.5.	<i>(Continuación)</i> Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 4 y 5 . . . . .	27
3.6.	<i>(Continuación)</i> Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de diemsión 5 y 10 . . . . .	28
3.7.	Índices para los intervalos de confianza de una Distribución Multinomial de dimensión 3 . . . . .	29
3.8.	Índices para los intervalos de confianza de una Distribución Multinomial de dimensión 3 y 4 . . . . .	30
3.9.	Índices para los intervalos de confianza de una Distribución Multinomial de dimensión 4, 5 y 10 . . . . .	31
3.10.	Índice Promedio . . . . .	37
4.1.	Resultados de los intervalos de confianza basados en la Distribución Multinomial aplicados a la base de datos de Muertos en 1996 en la ciudad de Medellín . . . . .	41

# Capítulo 1

## Introducción

Un problema común en el trabajo estadístico consiste en estimar los parámetros que ayudan a caracterizar una variable. El cálculo de Intervalos de Confianza para la estimación de parámetros permite hacer declaraciones sobre qué valores se pueden esperar sobre éstos. En la aplicación estadística, para el análisis de resultados, cada vez se prefiere más el uso de intervalos de confianza que las pruebas de hipótesis, debido a que el intervalo de confianza aporta información para la magnitud y la precisión de las estimaciones.

Los intervalos de confianza son prácticos y atractivos a la hora de presentar resultados, mientras que el valor  $p$  en las pruebas de hipótesis presentan una elaboración probabilística de interpretación más compleja.

La construcción de intervalos de confianza para los parámetros de la distribución multinomial es un problema que se presenta frecuentemente en el trabajo estadístico aplicado; un caso típico, es cuando se responden preguntas de selección múltiple.

Cabe anotar que los intervalos de confianza que presentan los textos básicos de estadística, construidos con base en el Teorema del Límite Central, tienen un desempeño bajo ya que pueden resultar intervalos con un nivel de confianza real por debajo del nivel de confianza nominal.

El objetivo de este trabajo es comparar las diferentes metodologías para calcular los intervalos de confianza de los parámetros de la distribución multinomial. Para cada uno de los intervalos se analiza el nivel de confianza, la longitud y el índice, teniendo en cuenta la varianza para los distintos valores de  $\pi$  y el tamaño de la muestra  $N$ .

En el capítulo 2 se presentan aspectos teóricos de los intervalos que van



a compararse. En el capítulo 3, se describe la metodología usada en la comparación de los intervalos incluyendo algunos resultados y conclusiones, en el capítulo 4 se presenta una aplicación del estudio comparativo a una base de datos real de las muertes en Medellín y finalmente en el capítulo 5 se incluyen algunas conclusiones y recomendaciones.

## 1.1. Planteamiento del problema

El problema de encontrar intervalos de confianza para los parámetros de la Distribución Multinomial es aplicado en muchas áreas. Surge naturalmente cuando se responden preguntas de selección múltiple. Siendo tan importante es extraño el poco énfasis al trabajo inferencial que realizan los textos de Estadística sobre el tema. Para la parte inferencial generalmente se recurre al Teorema del Límite Central, asumiendo que los tamaños muestrales son lo suficientemente grandes.

La determinación del tamaño muestral ha sido estudiada por muchos investigadores como Hurtubise (1969); Angers (1974); Angers (1979); Angers (1984); Angers (1989); Tortora (1978); Thompson (1987) y Bromaghin (1993).

En este trabajo se realiza una revisión teórica y un estudio comparativo de las metodologías propuestas en la literatura con el fin de establecer diferencias y recomendar la más apropiada de acuerdo a su buen desempeño.

# Capítulo 2

## Marco Teórico

En esta sección se presentan algunos conceptos básicos sobre la Distribución Multinomial y algunas conclusiones importantes de investigaciones realizadas sobre los intervalos de confianza para dicha distribución.

### 2.1. Modelo multinomial

Una generalización inmediata de la distribución binomial surge cuando cada ensayo tiene más de dos resultados posibles, las probabilidades de los resultados correspondientes son las mismas para cada ensayo, y los ensayos son todos independientes.

Considere el caso donde hay  $N$  ensayos independientes que permiten  $k$  resultados mutuamente excluyentes cuyas probabilidades respectivas son  $\pi_1, \pi_2, \dots, \pi_k$ , con  $\sum_{i=1}^k \pi_i = 1$ . Al referirse a los resultados como que son de la primera clase, la segunda clase y la  $k$ -ésima clase se está interesado en la probabilidad de obtener  $x_1$  resultados de la primera clase,  $x_2$  resultados de la segunda clase y  $x_k$  resultados de la  $k$ -ésima clase, con  $\sum_{i=1}^k x_i = N$ . Considere el caso  $k = 2$ . Cada resultado pertenece o bien al éxito o al fracaso y  $\pi_2 = 1 - \pi_1$ . Entonces  $x_1$  es el número de éxitos en  $N$  repeticiones independientes y  $x_2 = N - x_1$  es el número de fracasos, luego la probabilidad de obtener  $x_1$  éxitos y  $N - x_1$  fracasos en  $N$  repeticiones independientes viene dada por una distribución binomial,

$$\binom{N}{x_1} \pi_1^{x_1} (1 - \pi_1)^{N-x_1} = \frac{N!}{x_1!(N-x_1)!} \pi_1^{x_1} (1 - \pi_1)^{N-x_1} = \frac{N!}{x_1!x_2!} \pi_1^{x_1} \pi_2^{x_2} \quad (2.1)$$

Luego, las variables aleatorias  $X_1, X_2, \dots, X_k$  tienen una distribución multinomial y se conocen como variables aleatorias multinomiales si y sólo

si su distribución de probabilidad conjunta está dada por:

$$f(x_1, x_2, \dots, x_k; N, \pi_1, \pi_2, \dots, \pi_k) = \binom{N}{x_1, x_2, \dots, x_k} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k} \quad (2.2)$$

El nombre multinomial se deriva del hecho que para los valores de  $x_i$  las probabilidades son iguales a los términos correspondientes de la expansión multinomial de  $(\pi_1 + \pi_2 + \dots + \pi_k)^N$

## 2.2. Intervalos de confianza

Walpole (1992) presenta la definición de intervalo de confianza para una proporción o una media o una razón. Para la hipótesis  $\theta = \theta_0$ , donde  $\theta_0$  es cualquier valor concreto del parámetro desconocido,  $\theta$ . El conjunto de valores de  $\theta_0$ , tales que el nivel de significación es mayor o igual que  $\alpha$  se llama Intervalo de Confianza para  $\theta$  al  $100(1 - \alpha)\%$ . Chen (1990) define el intervalo de confianza para  $\theta \in \Theta$ , denotado por  $I_n$ , donde  $\gamma(\theta; I_n) = P_\theta(\theta \in I_n)$  es llamada la probabilidad de cobertura de  $I_n$ .

$$\inf_{\theta \in \Theta} \gamma(\theta; I_n)$$

es el coeficiente de confianza de  $I_n$  y

$$\lim_{n \rightarrow \infty} \inf_{\theta \in \Theta} \gamma(\theta; I_n)$$

es el coeficiente de confianza uniforme de la sucesión  $I_n$ .

### 2.2.1. Estimación del intervalo bayesiano

De manera similar al intervalo de confianza clásico, en el análisis bayesiano, se puede calcular un intervalo bayesiano  $(1 - \alpha)100\%$  empleando la distribución a posteriori. El intervalo  $a < \theta < b$  se llamará intervalo bayesiano  $(1 - \alpha)100\%$  para  $\theta$  si

$$\int_{-\infty}^a \pi(\theta|x)d\theta = \int_b^{\infty} \pi(\theta|x)d\theta = \frac{\alpha}{2} \quad (2.3)$$

En la aproximación bayesiana, la estimación por intervalos se define por una evaluación simple de las distribuciones a posteriori de los parámetros. Así, si  $\theta \in \Theta$  es una cantidad desconocida,  $C \in \Theta$  es una región de  $(1 - \alpha)100\%$  de credibilidad para  $\theta$  si  $P(\theta \in C|x) \geq 1 - \alpha$ . En este caso,  $1 - \alpha$  es llamado

el nivel de credibilidad. Si  $\theta$  es un escalar, la región  $C$  está dada usualmente por un intervalo  $[c_1, c_2]$  (Bernardo & Smith (2000)).

Para calcular los intervalos bayesianos se utiliza la distribución Dirichlet como distribución a priori conjugada de la distribución multinomial. Es decir, su función de densidad de probabilidad devuelve la credibilidad de que las probabilidades de  $k$  celdas son  $x_i$ , dado que cada celda ha sido observada  $\theta_i - 1$  veces, donde  $\theta$  es el parámetro de la distribución Dirichlet,  $\theta$  no negativo y real.

### 2.2.2. Intervalos basados en el método de Sison y Glaz

Sison & Glaz (1995) proponen dos formas de calcular intervalos de confianza simultáneos para los parámetros multinomiales. El primer método está basado en la aproximación para las probabilidades de la multinomial usando el algoritmo de Levin (1981). El segundo método utiliza la estructura de dependencia negativa inherente en la distribución multinomial y relaciona desigualdades de probabilidad introducidas por Glaz & Johnson (1984).

Sean  $x_1, \dots, x_k$  las frecuencias en una muestra de  $N$  observaciones de una distribución multinomial con probabilidades  $\pi_1, \dots, \pi_k$  donde  $\pi_i \geq 0$  y  $\sum_{i=1}^k \pi_i = 1$ .

Para  $i = 1, \dots, k$ . Sean  $V_i$  variables aleatorias independientes con distribución Poisson con media  $Np_i$ . Sea  $W$  la suma de las  $k$  observaciones independientes de una distribución Poisson truncada en el intervalo  $[b_i, a_i]$ , es decir

$W = \sum_{i=1}^k Y_i$  con  $E(W) = \sum_{i=1}^k \mu_i$  y varianza  $\sum_{i=1}^k \sigma_i^2$ , entonces

$$P(b_i \leq X_i \leq a_i, i = 1, \dots, k) \approx \frac{N!}{N^N e^{-N}} \left\{ \prod_{i=1}^k P(b_i \leq V_i \leq a_i) \right\} P(W = N)$$

$$= \frac{N!}{N^N e^{-N}} \left\{ \prod_{i=1}^k P(b_i \leq V_i \leq a_i) \right\} \times f_e \left( \frac{N - \sum_{i=1}^k \mu_i}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \right) \times \frac{1}{\sqrt{\sum_{i=1}^k \sigma_i^2}}$$

Patel & Read (1996) presentan la expansión Edgeworth

$$f_e(x) = \phi(x) \left[ 1 + \gamma_1 \left( \frac{H_3(x)}{3!} \right) + \gamma_2 \left( \frac{H_4(x)}{4!} \right) + 10\gamma_1^2 \left( \frac{H_6(x)}{6!} \right) + \dots \right]$$

donde

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

es la función de densidad de la distribución normal estándar, y

$$H_0(x) = 1$$

$$H_3(x) = x^3 - 3x$$

$$H_4(x) = x^4 - 6x^2 + 3$$

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15$$

son los polinomios de Tchebyshev-Hermite.

Luego

$$f_e(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \left\{ 1 + \frac{\gamma_1}{6}(x^3 - 3x) + \frac{\gamma_2}{24}(x^4 - 6x^2 + 3) + \frac{\gamma_1^2}{72}(x^6 - 15x^4 + 45x^2 - 15) \right\}$$

Si  $A_i = \{V_i/V_i \in [b_i, a_i]\}$  es el conjunto de eventos tales que  $b_i \leq V_i \leq a_i$ , entonces por el Teorema de Bayes

$$\begin{aligned} P(A_1, \dots, A_k | \sum_{i=1}^k v_i = N) &= \frac{P(A_1, \dots, A_k)}{P(\sum_{i=1}^k v_i = n)} \times P(\sum_{i=1}^k v_i = N | A_1, \dots, A_k) \\ &= \frac{\prod_{i=1}^k P(b_i \leq V_i \leq a_i)}{\frac{N^N e^{-N}}{N!}} \times P(W = N) \end{aligned}$$

Sean  $\mu_i$ ,  $\sigma_i^2$ ,  $\mu_{3,i}$  y  $\mu_{4,i}$  los cuatro momentos centrales de la Distribución Poisson Truncada, luego

$$\begin{aligned} \gamma_1 &= \frac{1}{\sqrt{k}} \frac{\left( \frac{1}{k} \sum_{i=1}^k \mu_{3,i} \right)}{\left( \frac{1}{k} \sum_{i=1}^k \sigma_i^2 \right)^{3/2}} \\ \gamma_2 &= \frac{1}{\sqrt{k}} \frac{\left( \frac{1}{k} \sum_{i=1}^k \mu_{4,i} - 3\sigma_i^2 \right)}{\left( \frac{1}{k} \sum_{i=1}^k \sigma_i^2 \right)^2} \end{aligned}$$

donde  $\gamma_1$  es el coeficiente de sesgo y  $\gamma_2$  es el coeficiente de kurtosis.

Sison & Glaz (1995) sugieren encontrar un entero  $c$  tal que  $\pi(c) \leq 1 - \alpha \leq \pi(c + 1)$ , teniendo en cuenta que la Distribución Multinomial es sesgada y usando un ajuste de interpolación para los intervalos de confianza simultáneos estos están dados por

$$\left( \hat{\pi}_i - \frac{c}{N} \leq \pi_i \leq \hat{\pi}_i + \frac{c}{N} + \frac{2\delta}{N} \right)$$

donde

$$\delta = \frac{(1 - \alpha) - \pi(c)}{\pi(c + 1) - \pi(c)}$$

### 2.2.3. Intervalo de Quesenberry y Hurst

Quesenberry & Hurst (1964) consideran cada celda versus las restantes como una distribución binomial y sugieren hacer un conjunto de estimaciones de intervalos de confianza binomiales para la proporción de las celdas individuales. El método está basado en la aproximación de la distribución chi cuadrada.

El intervalo para  $\pi_i$  está dado por

$$\frac{\chi_{k-1,1-\alpha}^2 + 2x_i \pm \sqrt{\chi_{k-1,1-\alpha}^2 (\chi_{k-1,1-\alpha}^2 + 4\frac{x_i}{N}(N - x_i))}}{2(N + \chi_{k-1,1-\alpha}^2)}$$

Una mejora se logra si se se trabaja con  $1 - \alpha/k$  en lugar de  $1 - \alpha$ .

### 2.2.4. Método basado en el Teorema del Límite Central

Si el tamaño muestral es lo suficientemente grande, podemos aplicar el teorema central del límite multivariable. Si  $\mathbf{n} = (n_1, n_2, \dots, n_k)^T$  es un vector aleatorio  $k$ -dimensional proveniente de una *MULTINOMIAL*( $\pi, N$ ), donde  $N = \sum n_j$  es fijo y  $\pi = (\pi_1, \pi_2, \dots, \pi_k)$ , con  $\sum \pi_j = 1$ . Entonces

$$\hat{\pi} = \frac{1}{N} \mathbf{n}$$

Este es el intervalo propuesto en la mayoría de textos básicos de estadística (Roussas (1973); Walpole (1992); Casella & Berger (2002); Meyer (1986); Canavos (1988)).

$$\left( \hat{\pi}_i - z_{\alpha/(2k)} \sqrt{\frac{\hat{\pi}_i(1 - \hat{\pi}_i)}{N}}, \hat{\pi}_i + z_{\alpha/(2k)} \sqrt{\frac{\hat{\pi}_i(1 - \hat{\pi}_i)}{N}} \right)$$

Se puede considerar la corrección por continuidad propuesta por Snedecor & Cochran (1980)

$$\left( \hat{\pi}_i - z_{\alpha/(2k)} \sqrt{\frac{\hat{\pi}_i(1 - \hat{\pi}_i)}{N}} - \frac{1}{2N}, \hat{\pi}_i + z_{\alpha/(2k)} \sqrt{\frac{\hat{\pi}_i(1 - \hat{\pi}_i)}{N}} + \frac{1}{2N} \right)$$

y el intervalo propuesto por Agresti & Caffo (2000)

$$\hat{\pi}_i \left( \frac{N}{N + z_{\alpha/2}^2} \right) + \frac{1}{2} \left( \frac{z_{\alpha/2}^2}{N + z_{\alpha/2}^2} \right) \pm z_{\alpha/2}^2 \sqrt{\left( \frac{1}{N + z_{\alpha/2}^2} \right) \left[ \hat{\pi}_i(1 - \hat{\pi}_i) \left( \frac{N}{N + z_{\alpha/2}^2} \right) + \frac{1}{4} \left( \frac{z_{\alpha/2}^2}{N + z_{\alpha/2}^2} \right) \right]}$$

### 2.2.5. Intervalos basados en la Razón de Verosimilitud Relativa

Kalbfleish (1985) presenta la metodología para construir intervalos de verosimilitud. Si  $L(\theta)$  es la función de verosimilitud, se define la *función de verosimilitud relativa* como

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}$$

El conjunto de valores de  $\theta$  para los cuales  $R(\theta) \geq p$  es llamado *intervalo de 100 %p de verosimilitud* para  $\theta$ . Se deben hallar las raíces que nos dan los límites del intervalo. Para el caso del parámetro de la Bernoulli,  $\pi$ , tenemos que un intervalo de confianza del 95 % se halla encontrando el par de raíces tal que

$$R(\pi_1, \pi_2, \dots, \pi_k) = \frac{L(\pi_1, \pi_2, \dots, \pi_k)}{L(\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_k)} \geq K(k, \alpha)$$

Esta desigualdad se resuelve numéricamente.

### 2.2.6. Método exacto basado en la $F$

Para construir este intervalo con un nivel  $(1 - \alpha)100\%$  de confianza para  $\pi$  se deben determinar los límites inferior  $L_I$  y superior  $L_S$ , tales que  $P(Y \geq y | \pi = L_I) = \alpha/2$  y  $P(Y \leq y | \pi = L_S) = \alpha/2$ . Leemis & Trivedi (1996) muestran dos procedimientos mediante los cuales se calculan  $L_I$  y  $L_S$  en términos de la distribución  $F$ . El intervalo “exacto” es:

$$\left( \frac{1}{1 + \frac{n-y+1}{yF_{2y,2(n-y+1),1-\alpha/2}}}, \frac{1}{1 + \frac{n-y}{(y+1)F_{2(y+1),2(n-y),\alpha/2}}} \right)$$

El método bootstrap proporciona una manera directa y sencilla para hallar intervalos simultáneos para los parámetros de la distribución multinomial. Para hallarlos se procede así:

1. A partir de la muestra estime los parámetros por máxima verosimilitud.

$$\hat{\pi}_i = \frac{n_i}{N} \quad i = 1, 2, \dots, k$$

2. Genere  $M$  muestras de tamaño  $N$  de una distribución multinomial con parámetros  $\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_k$ . Para cada muestra estime los parámetros  $\pi_1, \pi_2, \dots, \pi_k$ , se puede decir que para la muestra  $j$  los estimadores son  $\hat{\pi}_1^j, \hat{\pi}_2^j, \dots, \hat{\pi}_k^j$
3. Para cada  $\{\hat{\pi}_i^j\}_{j=1}^M$ , construya un histograma y calcule los percentiles  $.025/(k-1)$  y  $0.975/(k-1)$ , estos se denotan por  $\hat{\pi}_i^{0,025}$  y  $\hat{\pi}_i^{0,975}$

### 2.2.7. Intervalo de Goodman

Goodman (1965) modificó el procedimiento de Quesenberry & Hurst (1964) y propuso el siguiente intervalo de confianza para  $\pi_i$

$$\frac{A + 2x_i \mp \sqrt{A(A + 4x_i \frac{(N-x_i)}{N})}}{2(N + A)}$$

donde  $A = \chi_{(1-\alpha/k),1}^2$ .

El método de Goodman (1965) está basado en la aproximación de la normal para una proporción binomial y utiliza la desigualdad de Bonferroni para poner un límite en la probabilidad, estos intervalos son más pequeños que los encontrados por Quesenberry & Hurst (1964).

### 2.2.8. Intervalo de Bailey

Bailey (1980) presenta tres conjuntos de intervalos de confianza para las probabilidades de una Distribución Multinomial. Todos están basados en la



desigualdad de Bonferroni. El primero de estos métodos, originalmente propuesto por Goodman (1965) está basado en la aproximación de la normal para una proporción binomial, mientras que los otros dos requieren transformaciones de normalidad, una transformación angular en un caso y una raíz cuadrada en la otra. Los dos últimos intervalos propuestos por Bailey (1980) para  $\pi_i$  son:

$$(\sin(\sin^{-1})\sqrt{p'_i} \mp \chi^2/(4N + 2)^{1/2})^2 \quad (2.4)$$

donde  $p'_i = (x_i + 3/8)/(N + 3/4)$ , con una  $\chi^2_{(1-\alpha/k),1}$ . El otro intervalo está dado por

$$\frac{[\sqrt{p''_i} \mp \sqrt{C(C + 1 - p''_i)}]^2}{(C + 1)^2} \quad (2.5)$$

donde  $p''_i = (x_i + 3/8)/(N + 1/8)$  y  $C = \chi^2/4N$  con una  $\chi^2_{(1-\alpha/k),1}$ . Ambos intervalos necesitan modificaciones cuando  $x_i$  está cerca a cero o  $N$ .

El intervalo dado por 2.4 debe ser reemplazado por  $\pi_i^- = 0$  si  $x_i \leq (N + 3/4)\{\sin[\chi^2/(4N + 2)^{1/2}]\}^2 + 1/8$  y  $\pi_i^+ = 1$  si  $x_i \geq (N + 3/4)\{\sin[\pi/2 - \chi^2/(4N + 2)^{1/2}]\}^2 - 7/8$ .

Para el intervalo definido por 2.5, las condiciones son  $\pi_i^- = 0$  si  $x_i \leq (N + 1/8)C - 3/8$  y  $\pi_i^+ = 1$  si  $x_i \geq (N - 1/4)$ , es decir  $x_i = N$ .

### 2.2.9. Intervalo de Fitzpatrick y Scott

Fitzpatrick & Scott (1987) proponen un intervalo de la forma

$$\hat{\pi}_i - \frac{z_{\alpha/2}}{2\sqrt{N}} < \pi_i < \hat{\pi}_i + \frac{z_{\alpha/2}}{2\sqrt{N}}$$

donde  $\hat{\pi}_1, \dots, \hat{\pi}_k$  son las proporciones observadas en una muestra de tamaño  $N$  de una Distribución Multinomial con probabilidades  $\pi_1, \dots, \pi_k$  y  $\sum_{i=1}^k \pi_i = 1$ .

## 2.3. Entropía como medida de polarización en la multinomial

La entropía ha sido utilizada como una medida de incertidumbre Burrows (1989). Para una distribución multinomial con  $k$  categorías se define como

$$H = - \sum_{i=1}^k \pi_i \log(\pi_i).$$

Es bien conocido que la entropía es máxima para la distribución uniforme,  $H = \log(k)$ ; y si toda la masa de probabilidad está concentrada en una de las categorías se tiene  $H = 0$  ( $0 \leq H \leq \log(k)$ ). Si se tiene una muestra con  $N$  sujetos y se clasifican, se obtienen  $n_1, n_2, \dots, n_k$  frecuencias tal que  $\sum_{i=1}^k n_i = N$ , por lo tanto el estimador de la entropía será:

$$\hat{H} = - \sum_{i=1}^k \frac{n_i}{N} \log \left( \frac{n_i}{N} \right).$$

## 2.4. Estado del Arte

Quesenberry y Hurst (1964) presentan un método para obtener un conjunto de intervalos simultáneos para las probabilidades de una distribución multinomial. Un enfoque a este problema es considerar cada celda versus las restantes como una distribución binomial y hacer un conjunto de estimaciones de intervalos de confianza binomiales para la proporción de las celdas individuales. El método está basado en la aproximación de la distribución chi-cuadrada.

Goodman (1965) modificó el procedimiento de Quesenberry & Hurst (1964) presentando dos métodos de construir intervalos de confianza simultáneos. Ambos métodos de Goodman producen intervalos de confianza más pequeños que los de Quesenberry y Hurst (1964). El método de Goodman estuvo basado en la aproximación de la normal para una proporción binomial y utilizó la desigualdad de Bonferroni para poner un límite en la probabilidad de que todos los intervalos serían simultáneamente correctos.

Cabe anotar que el problema del tamaño muestral ha sido estudiado por muchos investigadores como Hurtubise (1969), Angers (1974, 1979, 1984, 1989) Tortora (1978) y Thompson (1987). Sus procedimientos estuvieron

basados en el enfoque de Goodman (1965).

Angers (1974) aplicó el método de Goodman al problema del tamaño de muestra para las proporciones multinomiales y presentó un método gráfico para seleccionar un tamaño de muestra basado en un conocimiento a priori de los valores del parámetro.

Tortora (1978) basó el tamaño de muestra para la distribución multinomial en el parámetro individual “caso peor”; el cual es el parámetro más cercano a 0.5 cuando los criterios de precisión son iguales para todos los parámetros.

Angers (1979) precisó que el método de Tortora era más conservador que necesario en algunos casos. Angers (1979, 1984) describe un procedimiento general para seleccionar el tamaño de muestra usando estimadores a priori de los valores del parámetro. Para intervalos de igual longitud, Angers (1984) da un resultado empírico sobre el “ caso peor” del vector de parámetros para niveles pequeños de  $\alpha$  basado en el método de Monte Carlo.

Thompson (1987) establece la forma del “ caso peor” para el vector de los parámetros de la multinomial cuando la longitud del intervalo se especifica para cada parámetro. Thompson (1987) presentó una formula para el tamaño de muestra bajo este “ caso peor” y una tabla que proporciona el tamaño de muestra, el nivel de significancia y la amplitud del intervalo. Aunque los procedimientos para la determinación del tamaño de muestra de Tortora (1978) y Thompson (1987) están basados en alguno de los métodos de Goodman, ellos tienen acercamientos diferentes al problema.

Bromaghin (1993) también estudió la determinación del tamaño muestral, comparó y resaltó las diferencias existentes entre los dos procedimientos de Tortora (1978) y Thompson (1987); también presentó un procedimiento empleando la metodología de Tortora (1978) y basado en el segundo método de Goodman.

El procedimiento propuesto por Bromaghin (1993) proporciona un tamaño muestral ligeramente más pequeño que el propuesto por Tortora (1978).

Sison & Glaz (1995) proponen dos formas de calcular intervalos de confianza simultáneos para los parámetros multinomiales. El primer método está basado en la aproximación de las probabilidades de la multinomial utilizando el algoritmo de Levin(1981), el cual consiste en una representación para la distribución acumulada de la multinomial. El segundo método utiliza la estructura de dependencia negativa en la distribución multinomial y las desigualdades relacionadas con las probabilidades; introducidas por Glaz & Johnson (1984).

May & Johnson (2000) proporcionan macros en SAS para calcular estos intervalos. Ellos argumentan que estos intervalos funcionan mejor que los otros métodos cuando el número de categorías es grande y el número de observaciones no lo es tanto y si no existen celdas que polaricen las probabilidades.

Correa & Sierra (2001) realizan un estudio comparativo de las metodologías reportadas en la literatura estadística para encontrar los intervalos de confianza para el parámetro de la distribución binomial, encontrando que el mejor método es el de la razón de verosimilitud seguido por el método exacto de la F.

Correa & Sierra (2003) revisan diferentes procedimientos de construcción de intervalos de confianza para la comparación de dos proporciones utilizando herramientas de simulación, encontrando que el método de Wald es el más deficiente, pero desafortunadamente es el que con mayor frecuencia utilizan los investigadores por su facilidad.

Cepeda et al. (2008) evalúan y comparan el comportamiento de diferentes metodologías empleadas para la obtención de intervalos de confianza e intervalos de credibilidad para una proporción, verifican conclusiones conocidas como el mal comportamiento del intervalo de Wald y recomiendan el intervalo score y bayesiano con distribución a priori uniforme, ya que presentan el mejor desempeño.

# Capítulo 3

## Estudio de Simulación

### 3.1. Metodología

Para la comparación via simulación, los factores que se tuvieron en cuenta así como sus respectivos niveles fueron:

- Tamaño de la muestra de la población (**N**): 30, 50, 100, 200 y 500.
- Dimensión de la distribución multinomial (*dim*): 3, 4, 5 y 10.  
Cuando se encontraban celdas con cero observaciones se reemplazaba  $n_i$  por 0.5
- Valores de  $\pi_i$ : Dimensión 3: 1/3, 1/3, 1/3; 0.3, 0.3, 0.4; 0.2, 0.3, 0.5; 0.1, 0.3, 0.6; 0.05, 0.3, 0.65; 0.05, 0.2, 0.75; 0.05, 0.1, 0.85; 0.05, 0.05, 0.9; 0.01, 0.01, 0.98.  
Dimensión 4: 1/4, 1/4, 1/4, 1/4; 0.4, 0.3, 0.2, 0.1; 0.7, 0.2, 0.05, 0.05; 0.7, 0.2, 0.095, 0.005;  
Dimensión 5: 1/5, 1/5, 1/5, 1/5, 1/5; 0.2, 0.2, 0.2, 0.15, 0.25; 0.2, 0.2, 0.2, 0.1, 0.3; 0.05, 0.05, 0.05, 0.05, 0.8; 0.2, 0.2, 0.2, 0.05, 0.35;  
Dimensión 10: 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10.
- Número de simulaciones (**Nsim**): 1000
- Nivel de confianza nominal ( $1 - \alpha$ ): 0.95

#### Variables de interés

- Nivel de confianza real (**level**): Es la proporción de intervalos que cubren el verdadero valor del parámetro.

- Longitud promedio de los intervalos (**vol**): Es el producto de las diferencias entre el límite superior e inferior de cada intervalo.
- Índice: Es la combinación entre el nivel de confianza real y la longitud promedio del intervalo.

Además se reportan los niveles de entropía para cada vector de  $(\pi)$ .

Las tablas 3.1 a 3.6 presentan el nivel de confianza real del intervalo y el volumen, las tablas 3.7, 3.8, 3.9 presentan el índice para cada intervalo encontrado, con cada combinación de  $\pi_i$ , para cada tamaño de muestra y para cada uno de los métodos. Los gráficos 3.1 a 3.5 muestran el nivel versus el volumen de cada intervalo para cada  $\pi_i$  y para cada tamaño de muestra. Finalmente se presenta un Cluster en la gráfica 3.6 para visualizar mejor la similitud que hay entre los métodos.

TABLA 3.1: Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 3

1/3,1/3,1/3										
Entropía Relativa=1										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.922	0.063111	0.955	0.032359	0.945	0.010645	0.932	0.003963	0.953	0.001000
IC.Bayes	0.865	0.029329	0.873	0.015095	0.855	0.005785	0.892	0.002136	0.889	0.000554
IC.QH	0.973	0.055747	0.945	0.028926	0.957	0.011182	0.961	0.004142	0.965	0.001078
IC.Goodman	0.966	0.052642	0.945	0.027229	0.957	0.010496	0.946	0.003880	0.952	0.001009
IC.F	0.991	0.082759	0.986	0.039200	0.968	0.013662	0.969	0.004684	0.967	0.001137
IC.Fitz	0.913	0.045821	0.898	0.021296	0.908	0.007529	0.901	0.002662	0.904	0.000673
IC.Boot	0.800	0.034173	0.861	0.016734	0.849	0.006045	0.881	0.002167	0.877	0.000555
IC.Bailey2	0.939	0.064661	0.937	0.029322	0.943	0.010913	0.951	0.003960	0.952	0.001018
IC.Bailey3	0.928	0.056391	0.946	0.028618	0.966	0.010792	0.949	0.003936	0.949	0.001015
IC.TCL	0.882	0.065545	0.957	0.031305	0.936	0.011280	0.959	0.004028	0.963	0.001025
IC.TCL.cor	0.879	0.065552	0.925	0.031342	0.949	0.011293	0.952	0.004029	0.960	0.001025
IC.TCL.II	0.913	0.031017	0.89	0.015656	0.838	0.005890	0.862	0.002155	0.88	0.000557
IC.RV	0.937	0.061026	0.970	0.030815	0.968	0.011560	0.962	0.004215	0.961	0.001086
.3,.3,.4										
Entropía Relativa=0.9911595										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.915	0.061984	0.920	0.031695	0.936	0.010632	0.941	0.003623	0.953	0.000944
IC.Bayes	0.884	0.028886	0.891	0.014828	0.872	0.005688	0.894	0.002098	0.868	0.000545
IC.QH	0.971	0.055108	0.952	0.028517	0.968	0.011006	0.955	0.004070	0.973	0.001059
IC.Goodman	0.972	0.051891	0.953	0.026843	0.961	0.010325	0.952	0.003817	0.969	0.000992
IC.F	0.988	0.081714	0.980	0.038577	0.975	0.013431	0.973	0.004601	0.968	0.001118
IC.Fitz	0.924	0.045821	0.880	0.021296	0.874	0.007529	0.909	0.002662	0.888	0.000673
IC.Boot	0.770	0.033197	0.806	0.016355	0.827	0.005939	0.847	0.002132	0.864	0.000545
IC.Bailey2	0.953	0.064460	0.967	0.028843	0.963	0.010715	0.953	0.003889	0.951	0.001000
IC.Bailey3	0.953	0.055682	0.957	0.028185	0.969	0.010596	0.964	0.003865	0.941	0.000997
IC.TCL	0.907	0.064403	0.931	0.030676	0.939	0.011068	0.95	0.003956	0.937	0.001005
IC.TCL.cor	0.907	0.064243	0.929	0.030737	0.934	0.011086	0.943	0.003952	0.955	0.001006
IC.TCL.II	0.829	0.030454	0.875	0.015366	0.846	0.005802	0.88	0.002119	0.883	0.000547
IC.RV	0.957	0.060064	0.964	0.030321	0.969	0.011351	0.977	0.004134	0.960	0.001067
.2,.3,.5										
Entropía Relativa=0.9372306										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.898	0.055050	0.920	0.026264	0.934	0.009795	0.935	0.003356	0.944	0.000875
IC.Bayes	0.883	0.026135	0.888	0.013371	0.888	0.005094	0.896	0.001874	0.885	0.000485
IC.QH	0.980	0.050560	0.950	0.025705	0.962	0.009875	0.957	0.003641	0.959	0.000945
IC.Goodman	0.981	0.047144	0.950	0.024344	0.941	0.009229	0.952	0.003416	0.950	0.000886
IC.F	0.987	0.075059	0.983	0.035062	0.969	0.012110	0.979	0.004147	0.965	0.001002
IC.Fitz	0.922	0.045821	0.902	0.021296	0.918	0.007529	0.918	0.002662	0.915	0.000673
IC.Boot	0.755	0.029332	0.831	0.014607	0.833	0.005297	0.862	0.001898	0.872	0.000485
IC.Bailey2	0.947	0.060300	0.971	0.026545	0.961	0.009575	0.951	0.003463	0.952	0.000891
IC.Bailey3	0.962	0.049699	0.956	0.025080	0.955	0.009419	0.964	0.003449	0.955	0.000888
IC.TCL	0.916	0.057294	0.935	0.027301	0.941	0.009869	0.949	0.003516	0.946	0.000894
IC.TCL.cor	0.914	0.057507	0.899	0.027277	0.941	0.009836	0.954	0.003521	0.953	0.000896
IC.TCL.II	0.886	0.027556	0.874	0.013841	0.855	0.005192	0.889	0.001893	0.885	0.000487
IC.RV	0.966	0.054187	0.961	0.026974	0.962	0.010150	0.96	0.003691	0.966	0.000950

TABLA 3.2: (Continuación) Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 3

.1.,.3.,.6										
Entropía Relativa=0.8173454										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.907	0.037500	0.893	0.018137	0.916	0.007534	0.942	0.002844	0.955	0.000735
IC.Bayes	0.890	0.020525	0.897	0.010113	0.886	0.003822	0.895	0.001390	0.877	0.000356
IC.QH	0.963	0.040280	0.966	0.019895	0.976	0.007512	0.967	0.002699	0.957	0.000696
IC.Goodman	0.958	0.038021	0.968	0.019051	0.967	0.007005	0.959	0.002544	0.969	0.000651
IC.F	0.993	0.060952	0.987	0.027852	0.979	0.009366	0.962	0.003148	0.964	0.000751
IC.Fitz	0.943	0.045766	0.912	0.021295	0.927	0.007529	0.918	0.002662	0.925	0.000673
IC.Boot	0.704	0.021055	0.763	0.010402	0.830	0.003834	0.854	0.001385	0.852	0.000355
IC.Bailey2	0.967	0.043691	0.965	0.021108	0.952	0.007266	0.972	0.002550	0.961	0.000654
IC.Bailey3	0.971	0.038119	0.975	0.018934	0.963	0.007014	0.961	0.002534	0.954	0.000656
IC.TCL	0.907	0.041551	0.938	0.019946	0.934	0.007258	0.956	0.002589	0.949	0.000659
IC.TCL.cor	0.734	0.040582	0.842	0.019862	0.913	0.007162	0.925	0.002574	0.956	0.000657
IC.TCL.II	0.902	0.021424	0.875	0.010665	0.867	0.003886	0.906	0.001405	0.9	0.000359
IC.RV	0.968	0.040605	0.960	0.019962	0.964	0.007476	0.964	0.002726	0.969	0.000701
.05.,.3.,.65										
Entropía Relativa=0.7199877										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.912	0.027726	0.921	0.013884	0.925	0.005592	0.936	0.002232	0.933	0.000649
IC.Bayes	0.890	0.016416	0.894	0.007699	0.903	0.002770	0.88	0.000992	0.884	0.000255
IC.QH	0.966	0.033251	0.978	0.015894	0.972	0.005660	0.968	0.001973	0.959	0.000500
IC.Goodman	0.960	0.031488	0.972	0.014640	0.963	0.005226	0.962	0.001867	0.952	0.000470
IC.F	0.988	0.051071	0.990	0.022197	0.985	0.007155	0.988	0.002339	0.976	0.000547
IC.Fitz	0.951	0.045575	0.944	0.021265	0.954	0.007529	0.947	0.002662	0.934	0.000673
IC.Boot	0.900	0.013820	0.858	0.006869	0.770	0.002635	0.829	0.000974	0.850	0.000250
IC.Bailey2	0.980	0.032856	0.970	0.014964	0.964	0.005420	0.962	0.001858	0.956	0.000466
IC.Bailey3	0.977	0.028806	0.980	0.013569	0.959	0.004964	0.969	0.001805	0.969	0.000464
IC.TCL	0.954	0.029726	0.886	0.013915	0.933	0.004932	0.944	0.001820	0.956	0.000463
IC.TCL.cor	0.959	0.030132	0.905	0.014066	0.862	0.005054	0.894	0.001817	0.938	0.000462
IC.TCL.II	0.872	0.017376	0.905	0.008114	0.877	0.002830	0.889	0.001004	0.896	0.000256
IC.RV	0.976	0.031064	0.967	0.014537	0.969	0.005301	0.967	0.001927	0.965	0.000498
.05.,.2.,.75										
Entropía Relativa=0.625731										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.897	0.019523	0.894	0.010326	0.898	0.004030	0.925	0.001590	0.934	0.000450
IC.Bayes	0.865	0.013678	0.884	0.006270	0.901	0.002247	0.878	0.000795	0.876	0.000203
IC.QH	0.967	0.027760	0.955	0.012961	0.961	0.004549	0.969	0.001585	0.965	0.000400
IC.Goodman	0.965	0.026200	0.958	0.012020	0.967	0.004199	0.958	0.001490	0.966	0.000372
IC.F	0.993	0.042246	0.982	0.018413	0.985	0.005774	0.981	0.001895	0.972	0.000438
IC.Fitz	0.976	0.045577	0.971	0.021272	0.969	0.007529	0.973	0.002662	0.969	0.000673
IC.Boot	0.899	0.011136	0.836	0.005486	0.791	0.002094	0.847	0.000774	0.872	0.000198
IC.Bailey2	0.967	0.029379	0.980	0.012158	0.966	0.004373	0.963	0.001473	0.955	0.000368
IC.Bailey3	0.983	0.024046	0.956	0.011102	0.972	0.004030	0.962	0.001430	0.957	0.000368
IC.TCL	0.966	0.023574	0.887	0.010957	0.918	0.004001	0.926	0.001439	0.946	0.000368
IC.TCL.cor	0.951	0.024338	0.861	0.010988	0.843	0.003965	0.91	0.001437	0.936	0.000366
IC.TCL.II	0.904	0.014204	0.893	0.006519	0.883	0.002280	0.912	0.000810	0.871	0.000203
IC.RV	0.981	0.025107	0.975	0.011945	0.974	0.004276	0.97	0.001523	0.963	0.000392



TABLA 3.3: (Continuación) Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 3

.05,.1,.85										
Entropía Relativa=0.4716734										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.943	0.011083	0.911	0.005460	0.911	0.002268	0.886	0.000855	0.904	0.000226
IC.Bayes	0.904	0.009465	0.895	0.004199	0.876	0.001448	0.902	0.000507	0.886	0.000126
IC.QH	0.973	0.020036	0.976	0.008852	0.963	0.002943	0.972	0.001011	0.961	0.000249
IC.Goodman	0.959	0.018661	0.975	0.008180	0.959	0.002761	0.96	0.000934	0.968	0.000234
IC.F	0.990	0.031169	0.994	0.012845	0.989	0.003938	0.982	0.001223	0.966	0.000278
IC.Fitz	0.988	0.045532	0.997	0.021275	0.990	0.007529	0.99	0.002662	0.996	0.000673
IC.Boot	0.808	0.006928	0.780	0.003347	0.784	0.001297	0.833	0.000478	0.854	0.000123
IC.Bailey2	0.978	0.019109	0.973	0.008582	0.964	0.002835	0.954	0.000922	0.964	0.000227
IC.Bailey3	0.986	0.015619	0.972	0.007230	0.967	0.002509	0.974	0.000912	0.970	0.000229
IC.TCL	0.929	0.014495	0.874	0.006747	0.929	0.002514	0.946	0.000890	0.935	0.000229
IC.TCL.cor	0.803	0.014923	0.813	0.006976	0.828	0.002486	0.895	0.000884	0.926	0.000227
IC.TCL.II	0.906	0.009859	0.923	0.004316	0.883	0.001455	0.899	0.000504	0.886	0.000127
IC.RV	0.983	0.016106	0.969	0.007293	0.961	0.002704	0.964	0.000954	0.966	0.000243
.05,.05,.9										
Entropía Relativa=0.3589962										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.940	0.005753	0.950	0.002857	0.916	0.001296	0.898	0.000488	0.905	0.000126
IC.Bayes	0.854	0.007170	0.899	0.002908	0.898	0.000923	0.885	0.000314	0.870	0.000079
IC.QH	0.971	0.015182	0.978	0.006239	0.972	0.001986	0.977	0.000642	0.962	0.000156
IC.Goodman	0.969	0.014279	0.976	0.005831	0.967	0.001804	0.97	0.000593	0.965	0.000144
IC.F	0.990	0.024022	0.995	0.009045	0.992	0.002673	0.982	0.000803	0.971	0.000178
IC.Fitz	0.999	0.045338	0.998	0.021248	0.999	0.007528	0.998	0.002662	1.000	0.000673
IC.Boot	0.935	0.004363	0.782	0.002013	0.726	0.000783	0.841	0.000292	0.824	0.000075
IC.Bailey2	0.993	0.012758	0.985	0.005739	0.974	0.001910	0.963	0.000589	0.970	0.000139
IC.Bailey3	0.987	0.010833	0.985	0.004653	0.974	0.001617	0.961	0.000560	0.957	0.000139
IC.TCL	1.000	0.009916	0.851	0.004228	0.905	0.001483	0.933	0.000558	0.944	0.000140
IC.TCL.cor	0.998	0.009589	0.820	0.004169	0.762	0.001504	0.881	0.000548	0.924	0.000138
IC.TCL.II	0.89	0.007116	0.927	0.002913	0.88	0.000942	0.921	0.000318	0.874	7.87E-05
IC.RV	0.989	0.011417	0.988	0.004778	0.976	0.001662	0.961	0.000591	0.951	0.000149
.01,.01,.98										
Entropía Relativa=0.1018576										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.999	0.001090	0.998	0.000373	0.983	0.000098	0.886	0.000032	0.830	0.000010
IC.Bayes	0.516	0.003671	0.730	0.001097	0.797	0.000214	0.866	0.000048	0.904	0.000021
IC.QH	0.924	0.008435	0.969	0.002597	0.965	0.000500	0.963	0.000111	0.971	0.000026
IC.Goodman	0.936	0.007739	0.970	0.002369	0.958	0.000458	0.962	0.000105	0.974	0.000032
IC.F	0.996	0.013666	0.995	0.003878	0.993	0.000736	0.986	0.000153	0.990	0.000037
IC.Fitz	1.000	0.044189	1.000	0.020920	1.000	0.007486	1	0.002659	1.000	0.000043
IC.Boot	0.995	0.000753	0.981	0.000246	0.971	0.000070	0.721	0.000025	0.743	0.000048
IC.Bailey2	1.000	0.004359	1.000	0.001252	0.998	0.000271	0.992	0.000078	0.970	0.000054
IC.Bailey3	0.979	0.004776	0.981	0.001404	0.985	0.000278	0.976	0.000071	0.965	0.000059
IC.TCL	1.000	0.002838	1.000	0.000787	1.000	0.000187	0.983	0.000053	0.913	0.000065
IC.TCL.cor	0.999	0.002803	1.000	0.000813	0.999	0.000192	0.758	0.000057	0.781	0.000070
IC.TCL.II	0.926	0.003537	0.841	0.001021	0.852	0.000200969	0.892	4.71E-05	0.906	9.09E-06
IC.RV	0.991	0.004416	0.994	0.001293	0.992	0.000271	0.987	0.000069	0.972	0.000015

TABLA 3.4: (Continuación) Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 4

1/4,1/4,1/4,1/4										
Entropía Relativa=1										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.964	0.022698	0.940	0.007402	0.945	0.002342	0.924	0.000506	0.949	0.000085
IC.Bayes	0.800	0.005933	0.887	0.002517	0.863	0.000719	0.837	0.000193	0.833	0.000032
IC.QH	0.989	0.025406	0.996	0.010672	0.980	0.003011	0.988	0.000803	0.986	0.000134
IC.Goodman	0.968	0.017188	0.956	0.007025	0.945	0.001956	0.953	0.000517	0.963	0.000086
IC.F	0.990	0.038696	0.982	0.013886	0.977	0.003251	0.98	0.000750	0.970	0.000109
IC.Fitz	0.868	0.016396	0.924	0.005903	0.902	0.001476	0.892	0.000369	0.900	0.000059
IC.Boot	0.809	0.007631	0.787	0.002958	0.776	0.000777	0.798	0.000199	0.837	0.000032
IC.Bailey2	0.939	0.030475	0.953	0.008080	0.973	0.002037	0.954	0.000527	0.948	0.000086
IC.Bailey3	0.944	0.017693	0.958	0.007237	0.967	0.001987	0.958	0.000521	0.973	0.000086
IC.TCL	0.840	0.021410	0.901	0.008151	0.953	0.002118	0.948	0.000539	0.949	0.000087
IC.TCL.cor	0.843	0.021544	0.913	0.008170	0.898	0.002113	0.934	0.000538	0.942	0.000087
IC.TCL.II	0.902	0.007143	0.846	0.002831	0.836	0.000764933	0.858	0.000199	0.816	3.27E-05
IC.RV	0.990	0.027868	0.971	0.011279	0.977	0.003112	0.979	0.000817	0.986	0.000135
0.4,0.3,0.2,0.1										
Entropía Relativa= 0.9232197										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.922	0.017326	0.917	0.006807601	0.933	0.002100071	0.953	0.000501925	0.945	8.490E-05
IC.Bayes	0.839	0.004827	0.861	0.001985987	0.825	0.000554647	0.825	0.000149116	0.817	2.470E-05
IC.QH	0.98	0.021597	0.97	0.008720749	0.972	0.00239144	0.987	0.000627803	0.976	1.035E-04
IC.Goodman	0.958	0.014407	0.959	0.00571862	0.96	0.001559884	0.944	0.000402799	0.957	6.598E-05
IC.F	0.993	0.033870	0.988	0.011664655	0.983	0.002649256	0.969	0.00059793	0.968	8.549E-05
IC.Fitz	0.937	0.016380	0.894	0.005902606	0.904	0.001475681	0.894	0.00036892	0.907	5.903E-05
IC.Boot	0.608	0.005539	0.717	0.002212258	0.762	0.000590577	0.801	0.000152288	0.817	2.470E-05
IC.Bailey2	0.956	0.019914	0.96	0.006847779	0.953	0.001686813	0.955	0.000404	0.961	6.614E-05
IC.Bailey3	0.965	0.013922	0.975	0.005639067	0.952	0.001530684	0.948	0.000399615	0.958	6.598E-05
IC.TCL	0.852	0.016225	0.896	0.006083978	0.945	0.001611759	0.945	0.000411759	0.954	6.659E-05
IC.TCL.cor	0.712	0.016244	0.79	0.006092215	0.86	0.001615582	0.922	0.000410117	0.925	6.638E-05
IC.TCL.II	0.862	0.005774	0.851	0.002252791	0.785	0.000595014	0.863	0.000153703	0.86	2.51E-05
IC.RV	0.987	0.021914	0.992	0.00881689	0.985	0.002419776	0.98	0.000629644	0.981	1.033E-04
0.7, 0.2,0.05,0.05										
Entropía Relativa=0.6283898										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.899	0.005110	0.905	0.00212682	0.902	0.000642333	0.926	0.000211061	0.946	4.293E-05
IC.Bayes	0.838	0.002232	0.868	0.000795633	0.876	0.000196714	0.843	5.08E-05	0.824	8.150E-06
IC.QH	0.965	0.011528	0.97	0.004047164	0.979	0.000954664	0.975	0.000224854	0.984	3.485E-05
IC.Goodman	0.95	0.007353	0.965	0.002482702	0.957	0.000591113	0.955	0.000140855	0.957	2.201E-05
IC.F	0.991	0.019312	0.983	0.005812346	0.991	0.001146042	0.981	0.000236408	0.988	3.132E-05
IC.Fitz	0.969	0.016171	0.948	0.005885526	0.955	0.001475519	0.955	0.00036892	0.955	5.903E-05
IC.Boot	0.877	0.001529	0.762	0.000620103	0.681	0.000179618	0.763	4.78E-05	0.786	7.933E-06
IC.Bailey2	0.985	0.007270	0.97	0.002408555	0.962	0.000639095	0.962	0.000151709	0.959	2.145E-05
IC.Bailey3	0.973	0.005619	0.972	0.001997933	0.956	0.000500743	0.971	0.000131074	0.965	2.151E-05
IC.TCL	0.944	0.005456	0.814	0.00194816	0.9	0.000504365	0.932	0.000131262	0.943	2.154E-05
IC.TCL.cor	0.928	0.005409	0.795	0.001892756	0.745	0.000512058	0.922	0.000131149	0.924	2.141E-05
IC.TCL.II	0.825	0.002654	0.857	0.000899684	0.858	0.000214405	0.867	5.22E-05	0.831	8.30E-06
IC.RV	0.99	0.008798	0.99	0.003187254	0.986	0.000808589	0.98	0.000208128	0.978	3.388E-05

TABLA 3.5: (Continuación) Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de dimensión 4 y 5

0.7,0.2,0.095,0.005										
Entropía Relativa=0.5927097										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.934	0.005246	0.932	0.002156262	0.941	0.000575504	0.924	0.000152502	0.943	2.49E-05
IC.Bayes	0.757	0.002076	0.872	0.00068814	0.792	0.00014357	0.791	2.96E-05	0.842	3.97E-06
IC.QH	0.985	0.011214	0.964	0.003704614	0.979	0.000759254	0.968	0.000153024	0.975	1.91E-05
IC.Goodman	0.956	0.006996	0.952	0.002258886	0.954	0.000464211	0.944	9.23E-05	0.952	1.17E-05
IC.F	0.995	0.018524	0.993	0.005425721	0.979	0.000954934	0.982	0.000167621	0.971	1.84E-05
IC.Fitz	0.979	0.015908	0.946	0.005810508	0.948	0.001466907	0.958	0.00036827	0.965	5.90E-05
IC.Boot	0.807	0.001142	0.832	0.000361452	0.856	8.29E-05	0.844	1.85E-05	0.789	3.11E-06
IC.Bailey2	0.974	0.006724	0.959	0.001904371	0.956	0.00036591	0.97	6.95E-05	0.962	1.01E-05
IC.Bailey3	0.956	0.005006	0.965	0.00162421	0.968	0.000330096	0.974	6.88E-05	0.961	9.79E-06
IC.TCL	0.908	0.004465	0.927	0.001370759	0.935	0.000277369	0.965	6.02E-05	0.887	8.96E-06
IC.TCL.cor	0.902	0.004520	0.899	0.001375021	0.943	0.000284999	0.94	5.96E-05	0.885	9.17E-06
IC.TCL.II	0.802	0.002441	0.866	0.00077824	0.789	0.000154215	0.824	3.12E-05	0.853	4.08E-06
IC.RV	0.989	0.008295	0.988	0.002616306	0.987	0.000540304	0.987	0.000111832	0.983	1.52E-05
1/5,1/5,1/5,1/5,1/5										
Entropía Relativa=1										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.954	0.006935	0.893	0.001610	0.907	0.000320	0.939	0.000055	0.958	6.60E-06
IC.Bayes	0.787	0.000997	0.865	0.000353	0.808	0.000076	0.792	0.000015	0.800	1.62E-06
IC.QH	0.985	0.011536	0.985	0.003836	0.991	0.000781	0.997	0.000149	0.983	1.58E-05
IC.Goodman	0.949	0.005187	0.953	0.001665	0.962	0.000329	0.960	0.000062	0.950	6.48E-06
IC.F	0.994	0.018969	0.999	0.005037	0.992	0.000772	0.979	0.000116	0.974	9.82E-06
IC.Fitz	0.943	0.005866	0.906	0.001636	0.918	0.000289	0.939	0.000051	0.934	5.20E-06
IC.Boot	0.582	0.001425	0.683	0.000448	0.692	0.000086	0.747	0.000016	0.764	1.63E-06
IC.Bailey2	0.913	0.010346	0.950	0.002565	0.956	0.000339	0.949	0.000063	0.941	6.51E-06
IC.Bailey3	0.930	0.004883	0.964	0.001614	0.953	0.000325	0.958	0.000061	0.963	6.47E-06
IC.TCL	0.780	0.006190	0.916	0.001887	0.931	0.000352	0.939	0.000064	0.952	6.57E-06
IC.TCL.cor	0.752	0.006143	0.770	0.001888	0.866	0.000352	0.927	0.000064	0.949	6.57E-06
IC.TCL.II	0.86	0.001448	0.808	0.000449	0.754	8.65E-05	0.811	1.60E-05	0.793	1.66E-06
IC.RV	0.994	0.011844	0.988	0.003908	0.990	0.000790	0.989	0.000150	0.993	1.58E-05
0.2,0.2,0.2,0.15,0.25										
Entropía Relativa=0.9674888										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.959	0.006641	0.918	0.001556	0.917	0.000318	0.935	0.000064	0.952	6.60E-06
IC.Bayes	0.779	0.000964	0.832	0.000342	0.817	0.000074	0.805	0.000015	0.792	1.57E-06
IC.QH	0.983	0.011280	0.988	0.003743	0.990	0.000760	0.985	0.000144	0.993	1.53E-05
IC.Goodman	0.946	0.005064	0.965	0.001622	0.973	0.000320	0.950	0.000060	0.963	6.28E-06
IC.F	0.996	0.018738	0.994	0.004933	0.992	0.000754	0.986	0.000113	0.973	9.52E-06
IC.Fitz	0.914	0.005864	0.921	0.001636	0.922	0.000289	0.919	0.000051	0.920	5.20E-06
IC.Boot	0.662	0.001378	0.698	0.000430	0.694	0.000083	0.756	0.000015	0.763	1.58E-06
IC.Bailey2	0.941	0.009852	0.945	0.002459	0.948	0.000332	0.947	0.000060	0.963	6.30E-06
IC.Bailey3	0.938	0.004716	0.953	0.001570	0.965	0.000315	0.940	0.000059	0.952	6.26E-06
IC.TCL	0.750	0.005934	0.862	0.001800	0.910	0.000340	0.939	0.000062	0.942	6.35E-06
IC.TCL.cor	0.777	0.006019	0.793	0.001811	0.872	0.000341	0.938	0.000062	0.935	6.36E-06
IC.TCL.II	0.849	0.001404	0.796	0.000437	0.738	8.36E-05	0.807	1.55E-05	0.777	1.61E-06
IC.RV	0.994	0.011507	0.988	0.003809	0.987	0.000762	0.991	0.000145	0.985	1.53E-05

TABLA 3.6: (Continuación) Resultados de las simulaciones de los intervalos de confianza basados en la Distribución Multinomial de diemnsión 5 y 10

0.2,0.2,0.2,0.1,0.3										
Entropía Relativa=0.9674888										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.933	0.005952	0.909	0.001399	0.939	0.000304	0.953	0.000075	0.937	6.60E-06
IC.Bayes	0.813	0.000896	0.798	0.000309	0.819	0.000066	0.787	0.000013	0.780	1.40E-06
IC.QH	0.989	0.010670	0.990	0.003474	0.984	0.000693	0.985	0.000130	0.989	1.36E-05
IC.Goodman	0.961	0.004710	0.971	0.001493	0.952	0.000288	0.947	0.000054	0.964	5.59E-06
IC.F	0.992	0.017717	0.997	0.004642	0.991	0.000697	0.986	0.000103	0.976	8.62E-06
IC.Fitz	0.943	0.005859	0.896	0.001636	0.923	0.000289	0.925	0.000051	0.923	5.20E-06
IC.Boot	0.515	0.001196	0.638	0.000377	0.700	0.000073	0.744	0.000014	0.762	1.40E-06
IC.Bailey2	0.952	0.008407	0.936	0.002123	0.956	0.000321	0.952	0.000054	0.966	5.59E-06
IC.Bailey3	0.950	0.004263	0.961	0.001385	0.961	0.000281	0.954	0.000053	0.957	5.57E-06
IC.TCL	0.806	0.005273	0.882	0.001607	0.934	0.000301	0.919	0.000055	0.938	5.65E-06
IC.TCL.cor	0.794	0.005263	0.806	0.001597	0.886	0.000301	0.947	0.000055	0.938	5.65E-06
IC.TCL.II	0.812	0.001282	0.804	0.000394	0.735	7.52E-05	0.806	1.38E-05	0.779	1.43E-06
IC.RV	0.991	0.010540	0.992	0.003428	0.992	0.000683	0.993	0.000129	0.988	1.36E-05
0.05,0.05,0.05,0.05,0.8										
Entropía Relativa=0.4831881										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.930	0.000490	0.903	0.000154	0.940	3.84E-05	0.926	1.027E-05	0.940	1.34E-06
IC.Bayes	0.782	0.000197	0.810	0.000051	0.833	8.28E-06	0.766	1.493E-06	0.802	1.47E-07
IC.QH	0.984	0.003496	0.982	0.000835	0.987	1.15E-04	0.984	1.757E-05	0.995	1.55E-06
IC.Goodman	0.940	0.001273	0.946	0.000292	0.950	4.34E-05	0.932	6.786E-06	0.952	6.23E-07
IC.F	0.999	0.006896	0.989	0.001404	0.993	1.52E-04	0.984	1.779E-05	0.976	1.21E-06
IC.Fitz	0.991	0.005653	0.983	0.001623	0.983	2.89E-04	0.980	5.110E-05	0.988	5.20E-06
IC.Boot	0.945	0.000088	0.680	0.000030	0.563	6.72E-06	0.697	1.315E-06	0.745	1.38E-07
IC.Bailey2	0.998	0.000938	0.996	0.000233	0.947	4.71E-05	0.938	8.168E-06	0.956	5.79E-07
IC.Bailey3	0.980	0.000702	0.984	0.000183	0.951	3.10E-05	0.940	5.611E-06	0.948	5.79E-07
IC.TCL	0.994	0.000590	0.730	0.000155	0.875	2.94E-05	0.908	5.626E-06	0.912	5.70E-07
IC.TCL.cor	0.993	0.000588	0.733	0.000153	0.839	2.96E-05	0.917	5.492E-06	0.921	5.78E-07
IC.TCL.II	0.768	0.000276	0.831	6.11E-05	0.838	9.54E-06	0.85	1.580E-06	0.79	1.53E-07
IC.RV	0.997	0.001934	0.999	0.000475	0.988	7.92E-05	0.989	1.423E-05	0.993	1.42E-06
0.2,0.2,0.2,0.05,0.35										
Entropía Relativa=0.9213696										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.958	0.004995	0.922	0.001179	0.910	0.000272	0.947	0.000063	0.937	6.62E-06
IC.Bayes	0.792	0.000758	0.834	0.000253	0.786	0.000051	0.822	0.000010	0.785	1.06E-06
IC.QH	0.985	0.009642	0.982	0.003005	0.990	0.000565	0.989	0.000103	0.994	1.05E-05
IC.Goodman	0.957	0.004163	0.967	0.001245	0.955	0.000232	0.949	0.000042	0.955	4.30E-06
IC.F	0.997	0.016319	0.992	0.004100	0.987	0.000591	0.986	0.000085	0.967	6.83E-06
IC.Fitz	0.927	0.005815	0.907	0.001633	0.892	0.000289	0.897	0.000051	0.911	5.20E-06
IC.Boot	0.671	0.000851	0.703	0.000271	0.665	0.000054	0.731	0.000010	0.743	1.05E-06
IC.Bailey2	0.947	0.006406	0.958	0.001534	0.958	0.000246	0.952	0.000045	0.958	4.25E-06
IC.Bailey3	0.953	0.003497	0.964	0.001080	0.957	0.000215	0.958	0.000040	0.966	4.22E-06
IC.TCL	0.822	0.003993	0.837	0.001186	0.917	0.000227	0.936	0.000041	0.938	4.27E-06
IC.TCL.cor	0.853	0.004048	0.766	0.001193	0.878	0.000223	0.932	0.000041	0.934	4.26E-06
IC.TCL.II	0.793	0.001102	0.806	0.000322	0.798	5.86E-05	0.816	1.06E-05	0.787	1.09E-06
IC.RV	0.994	0.008788	0.996	0.002700	0.992	0.000524	0.992	0.000098	0.992	1.03E-05
1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10										
Entropía Relativa=1										
Método	N=30		N=50		N=100		N=200		N=500	
	level	vol	level	vol	level	vol	level	vol	level	vol
IC.SG	0.961	1.3369E-06	0.970	1.921E-07	0.943	2.325E-08	0.938	6.122E-10	0.947	6.43E-12
IC.Bayes	0.908	2.2476E-08	0.670	3.706E-09	0.649	2.259E-10	0.614	1.042E-11	0.623	1.36E-13
IC.QH	0.993	3.6798E-04	0.999	3.439E-05	0.998	1.073E-06	0.999	3.085E-08	1.000	2.90E-10
IC.Goodman	0.913	8.4047E-06	0.903	6.487E-07	0.960	1.965E-08	0.922	5.944E-10	0.951	5.97E-12
IC.F	0.999	1.2909E-03	0.997	7.740E-05	0.989	1.129E-06	0.994	1.456E-08	0.990	5.32E-11
IC.Fitz	0.986	3.3310E-05	0.964	2.669E-06	0.981	8.365E-08	0.984	2.614E-09	0.986	2.68E-11
IC.Boot	0.063	4.8477E-08	0.215	7.364E-09	0.454	3.457E-10	0.459	1.296E-11	0.541	1.45E-13
IC.Bailey2	1.000	5.1161E-06	0.997	7.669E-07	0.966	7.533E-08	0.950	6.334E-10	0.949	5.59E-12
IC.Bailey3	0.996	2.7314E-06	0.989	3.025E-07	0.964	1.294E-08	0.949	4.837E-10	0.953	5.47E-12
IC.TCL	0.623	3.3327E-06	0.705	3.544E-07	0.777	1.438E-08	0.919	5.096E-10	0.911	5.62E-12
IC.TCL.cor	0.635	3.2678E-06	0.687	3.502E-07	0.766	1.446E-08	0.839	5.109E-10	0.925	5.61E-12
IC.TCL.II	0.74	1.8017E-07	0.702	1.419E-08	0.537	4.744E-10	0.645	1.521E-11	0.669	1.59E-13
IC.RV	1	1.8480E-04	1	1.672E-05	1	6.37E-07	0.999	2.257E-08	0.999	2.51E-10

Tabla 3.7: Índices para los intervalos de confianza de una Distribución Multinomial de dimensión 3

N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	TCL.II	RV
<b>1/3,1/3,1/3</b>													
30	0.9093	0.8838	0.9671	0.9633	0.9568	0.9170	0.8133	0.9245	0.9218	0.8676	0.8646	0.9312	0.9261
50	0.9727	0.9051	0.9660	0.9677	0.9972	0.9251	0.8911	0.9374	0.9673	0.9758	0.9432	0.9222	0.9896
100	0.9841	0.8948	0.9961	0.9968	1.0050	0.9486	0.8883	0.9818	1.0059	0.9741	0.9877	0.8769	1.0072
200	0.9772	0.9369	1.0074	0.9919	1.0152	0.9459	0.9254	0.9971	0.9950	1.0054	0.9981	0.9054	1.0084
500	1.0022	0.9353	1.0147	1.0011	1.0167	0.9509	0.9226	1.0011	0.9979	1.0126	1.0095	0.9258	1.0105
<b>.3,.3,.4</b>													
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	TCL.II	RV
30	0.9035	0.9036	0.9658	0.9701	0.9550	0.9281	0.7836	0.9385	0.9473	0.8932	0.8934	0.8461	0.9469
50	0.9377	0.9240	0.9735	0.9762	0.9918	0.9066	0.8345	0.9885	0.9790	0.9499	0.9478	0.9069	0.9840
100	0.9748	0.9127	1.0077	1.0011	1.0125	0.9131	0.8654	1.0028	1.0092	0.9775	0.9723	0.8854	1.0084
200	0.9869	0.9391	1.0012	0.9983	1.0195	0.9543	0.8897	0.9993	1.0108	0.9960	0.9887	0.9244	1.0242
500	1.0022	0.9132	1.0231	1.0190	1.0178	0.9341	0.9090	1.0001	0.9895	0.9853	1.0043	0.929	1.0094
<b>.2,.3,.5</b>													
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	TCL.II	RV
30	0.8932	0.9052	0.9794	0.9839	0.9610	0.9261	0.7714	0.9367	0.9623	0.9090	0.9068	0.9069	0.9617
50	0.9430	0.9222	0.9743	0.9757	0.9985	0.9293	0.8620	0.9950	0.9811	0.9573	0.9205	0.9073	0.9843
100	0.9735	0.9300	1.0026	0.9814	1.0076	0.9590	0.8722	1.0019	0.9958	0.9808	0.9808	0.8953	1.0024
200	0.9809	0.9414	1.0037	0.9987	1.0263	0.9637	0.9056	0.9976	1.0112	0.9954	1.0007	0.934	1.0068
500	0.9928	0.9311	1.0085	0.9991	1.0148	0.9625	0.9174	1.0012	1.0044	0.9949	1.0023	0.9311	1.0159
<b>.1,.3,.6</b>													
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	TCL.II	RV
30	0.9189	0.9176	0.9729	0.9701	0.9816	0.9472	0.7254	0.9734	0.9831	0.9151	0.7413	0.9291	0.9776
50	0.9230	0.9347	0.9966	0.9995	1.0100	0.9396	0.7948	0.9943	1.0069	0.9677	0.8687	0.9112	0.9904
100	0.9569	0.9291	1.0197	1.0108	1.0209	0.9684	0.8703	0.9948	1.0066	0.9760	0.9542	0.9091	1.0072
200	0.9888	0.9408	1.0151	1.0069	1.0094	0.9637	0.8977	1.0205	1.0090	1.0037	0.9712	0.9523	1.0120
500	1.0045	0.9228	1.0067	1.0193	1.0140	0.9730	0.8965	1.0109	1.0036	0.9983	1.0057	0.947	1.0193
<b>.05,.3,.65</b>													
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	TCL.II	RV
30	0.9334	0.9215	0.9830	0.9787	0.9869	0.9554	0.9343	0.9977	0.9988	0.9744	0.9791	0.9019	0.9955
50	0.9560	0.9338	1.0131	1.0082	1.0190	0.9726	0.8970	1.0058	1.0176	0.9197	0.9392	0.9449	1.0031
100	0.9682	0.9479	1.0174	1.0084	1.0294	0.9967	0.8084	1.0092	1.0045	0.9773	0.9028	0.9205	1.0146
200	0.9831	0.9254	1.0169	1.0107	1.0376	0.9942	0.8718	1.0108	1.0182	0.9919	0.9393	0.9348	1.0159
500	0.9815	0.9303	1.0090	1.0016	1.0268	0.9825	0.8945	1.0058	1.0195	1.0058	0.9869	0.9429	1.0153
<b>.05,.2,.75</b>													
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	TCL.II	RV
30	0.9258	0.8981	0.9896	0.9892	1.0011	0.9805	0.9358	0.9880	0.9951	0.9929	0.9767	0.9381	1.0087
50	0.9313	0.9247	0.9922	0.9963	1.0147	1.0004	0.8752	1.0190	0.9951	0.9235	0.8964	0.9339	1.0141
100	0.9415	0.9463	1.0070	1.0136	1.0309	1.0123	0.8309	1.0124	1.0190	0.9624	0.8839	0.9274	1.0209
200	0.9721	0.9235	1.0184	1.0069	1.0307	1.0215	0.8909	1.0122	1.0112	0.9733	0.9565	0.9592	1.0195
500	0.9827	0.9219	1.0154	1.0165	1.0227	1.0193	0.9177	1.0049	1.0070	0.9954	0.9849	0.9167	1.0133

TABLA 3.8: Índices para los intervalos de confianza de una Distribución Multinomial de dimensión 3 y 4

<b>.05, .1, .85</b>												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9816	0.9426	1.0037	0.9906	1.0096	0.9926	0.8446	1.0098	1.0217	0.9637	0.8326	0.9443
50	0.9537	0.9381	1.0183	1.0179	1.0329	1.0271	0.8183	1.0154	1.0158	0.9138	0.8498	0.9674
100	0.9568	0.9208	1.0107	1.0067	1.0370	1.0343	0.8242	1.0119	1.0153	0.9754	0.8694	0.9281
200	0.9318	0.9490	1.0221	1.0096	1.0324	1.0393	0.8764	1.0033	1.0243	0.9949	0.9413	0.9458
500	0.9514	0.9325	1.0113	1.0187	1.0166	1.0477	0.8988	1.0145	1.0208	0.9840	0.9745	0.9325
<b>.05, .05, .9</b>												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9838	0.8925	1.0066	1.0054	1.0171	1.0039	0.9799	1.0319	1.0277	1.0422	1.0405	0.9302
50	0.9971	0.9436	1.0231	1.0214	1.0379	1.0282	0.8215	1.0309	1.0320	0.8920	0.8596	0.9729
100	0.9630	0.9444	1.0211	1.0161	1.0414	1.0437	0.7636	1.0233	1.0236	0.9512	0.8009	0.9254
200	0.9448	0.9313	1.0278	1.0204	1.0329	1.0477	0.8850	1.0131	1.0110	0.9816	0.9269	0.9692
500	0.9525	0.9157	1.0125	1.0156	1.0219	1.0519	0.8673	1.0209	1.0072	0.9935	0.9725	0.9199
<b>.01, .01, .98</b>												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	1.0504	0.5412	0.9644	0.9776	1.0341	1.0061	1.0466	1.0480	1.0256	1.0496	1.0486	0.9713
50	1.0501	0.7676	1.0174	1.0186	1.0433	1.0306	1.0324	1.0513	1.0312	1.0518	1.0514	1.0450
100	1.0346	0.8388	1.0153	1.0080	1.0445	1.0448	1.0220	1.0502	1.0366	1.0524	1.0514	1.0439
200	0.9326	0.9115	1.0136	1.0125	1.0377	1.0498	0.7589	1.0441	1.0273	1.0347	0.7978	0.9389
500	0.8737	0.9516	1.0221	1.0252	1.0421	1.0526	0.7821	1.0210	1.0157	0.9610	0.8220	0.9537
<b>1/4, 1/4, 1/4</b>												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9917	0.8371	1.0146	1.0014	1.0018	0.8987	0.8451	0.9583	0.9761	0.8653	0.8683	0.9427
50	0.9821	0.9313	1.0372	0.9992	1.0193	0.9669	0.8260	0.9951	1.0011	0.9407	0.9532	0.888
100	0.9924	0.9078	1.0285	0.9928	1.0251	0.9481	0.8162	1.0221	1.0159	1.0010	0.9433	0.8793
200	0.9721	0.8809	1.0392	1.0026	1.0308	0.9386	0.8398	1.0037	1.0079	0.9974	0.9826	0.903
500	0.9989	0.8768	1.0378	1.0136	1.0209	0.9473	0.8810	0.9978	1.0241	0.9989	0.9915	0.8589
<b>0.4, 0.3, 0.2, 0.1</b>												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9537	0.8789	1.0093	0.9939	1.0099	0.9702	0.6365	0.9863	1.0016	0.8823	0.7373	0.9021
50	0.9587	0.9045	1.0121	1.0037	1.0279	0.9355	0.7531	1.0036	1.0205	0.9374	0.8265	0.8938
100	0.9800	0.8679	1.0207	1.0090	1.0320	0.9502	0.8016	1.0015	1.0006	0.9931	0.9038	0.8258
200	1.0027	0.8683	1.0383	0.9933	1.0194	0.9407	0.8430	1.0049	0.9975	0.9943	0.9701	0.9083
500	0.9947	0.8600	1.0273	1.0073	1.0189	0.9547	0.8600	1.0115	1.0084	1.0041	0.9736	0.9052
<b>0.7, 0.2, 0.05, 0.05</b>												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9415	0.8801	1.0041	0.9926	1.0230	1.0035	0.9217	1.0293	1.0185	0.9883	0.9716	1.0329
50	0.9506	0.9130	1.0169	1.0133	1.0287	0.9920	0.8016	1.0186	1.0211	0.8552	0.8353	0.9013
100	0.9489	0.9219	1.0295	1.0068	1.0420	1.0038	0.7167	1.0120	1.0058	0.9469	0.7838	0.903
200	0.9745	0.8873	1.0261	1.0051	1.0324	1.0049	0.8031	1.0125	1.0220	0.9809	0.9704	0.9126
500	0.9957	0.8674	1.0358	1.0073	1.0400	1.0073	0.8274	1.0095	1.0158	0.9926	0.9726	1.0294

Tabla 3.9: Índices para los intervalos de confianza de una Distribución Multinomial de dimensión 4, 5 y 10

0.7, 0.2, 0.095, 0.005												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9780	0.7952	1.0252	0.9993	1.0280	1.0141	0.8485	1.0184	1.0013	0.9515	0.9452	1.0324
50	0.9789	0.9173	1.0110	0.9998	1.0396	0.9900	0.8755	1.0076	1.0141	0.9745	0.9450	1.0373
100	0.9900	0.8336	1.0297	1.0037	1.0295	0.9964	0.9010	1.0059	1.0186	0.9839	0.9923	1.0384
200	0.9725	0.8326	1.0188	0.9936	1.0335	1.0080	0.8884	1.0210	1.0252	1.0157	0.9873	1.0388
500	0.9926	0.8863	1.0263	1.0021	1.0221	1.0157	0.8305	1.0126	1.0116	0.9337	0.9316	1.0347
1/5, 1/5, 1/5, 1/5, 1/5												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9972	0.8276	1.0249	0.9938	1.0265	0.9868	0.6118	0.9511	0.9742	0.8160	0.7867	1.0339
50	0.9385	0.9102	1.0329	1.0015	1.0463	0.9521	0.7186	0.9974	1.0131	0.9624	0.8090	1.0359
100	0.9544	0.8505	1.0423	1.0123	1.0434	0.9660	0.7284	1.0060	1.0028	0.9797	0.9113	1.0413
200	0.9884	0.8337	1.0493	1.0105	1.0304	0.9884	0.7863	0.9989	1.0084	0.9884	0.9757	1.0409
500	1.0084	0.8421	1.0347	1.0000	1.0253	0.9832	0.8042	0.9905	1.0137	1.0021	0.9989	1.0452
0.2, 0.2, 0.2, 0.15, 0.25												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	1.0028	0.8192	1.0231	0.9907	1.0288	0.9565	0.6959	0.9808	0.9827	0.7848	0.8130	1.0343
50	0.9648	0.8755	1.0361	1.0141	1.0412	0.9679	0.7344	0.9923	1.0016	0.9057	0.8332	1.0360
100	0.9650	0.8599	1.0413	1.0239	1.0434	0.9702	0.7305	0.9976	1.0155	0.9576	0.9176	1.0382
200	0.9841	0.8474	1.0367	0.9999	1.0378	0.9873	0.7958	0.9968	0.9804	0.9884	0.9873	1.0430
500	1.0021	0.8337	1.0452	1.0137	1.0242	0.9684	0.8032	1.0137	1.0021	0.9916	0.9842	1.0368
0.2, 0.2, 0.2, 0.1, 0.3												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9763	0.8550	1.0299	1.0068	1.0257	0.9868	0.5415	0.9937	0.9957	0.8439	0.8314	1.0322
50	0.9555	0.8397	1.0385	1.0206	1.0446	0.9416	0.6713	0.9832	1.0102	0.9269	0.8471	1.0406
100	0.9881	0.8620	1.0351	1.0018	1.0424	0.9713	0.7368	1.0060	1.0113	0.9829	0.9324	1.0435
200	1.0031	0.8284	1.0367	0.9968	1.0378	0.9736	0.7831	1.0021	1.0042	0.9673	0.9968	1.0451
500	0.9863	0.8211	1.0410	1.0147	1.0274	0.9716	0.8021	1.0168	1.0074	0.9874	0.9874	1.0400
0.2, 0.2, 0.2, 0.05, 0.35												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	1.0034	0.8331	1.0268	1.0032	1.0323	0.9701	0.7057	0.9905	0.9996	0.8618	0.8943	1.0371
50	0.9694	0.8777	1.0306	1.0166	1.0399	0.9532	0.7398	1.0069	1.0136	0.8800	0.8054	1.0456
100	0.9576	0.8273	1.0415	1.0050	1.0383	0.9387	0.7000	1.0082	1.0072	0.9650	0.9240	1.0437
200	0.9968	0.8653	1.0409	0.9989	1.0378	0.9442	0.7695	1.0021	1.0084	0.9852	0.9810	1.0441
500	0.9863	0.8263	1.0463	1.0053	1.0179	0.9589	0.7821	1.0084	1.0168	0.9874	0.9832	1.0442
0.05, 0.05, 0.05, 0.05, 0.8												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	0.9785	0.8230	1.0322	0.9882	1.0443	1.0373	0.9946	1.0495	1.0309	1.0457	1.0446	1.0474
50	0.9504	0.8526	1.0328	0.9955	1.0396	1.0331	0.7158	1.0482	1.0356	0.7683	0.7715	1.0511
100	0.9894	0.8768	1.0388	1.0000	1.0451	1.0344	0.5926	0.9968	1.0010	0.9210	0.8831	1.0399
200	0.9747	0.8063	1.0358	0.9810	1.0358	1.0315	0.7337	0.9874	0.9895	0.9558	0.9653	1.0410
500	0.9895	0.8442	1.0474	1.0021	1.0274	1.0400	0.7842	1.0063	0.9979	0.9600	0.9695	1.0453
1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10												
N	SG	Bayes	QH	Goodm	F	Fitz	Boot	Bailey2	Bailey3	TCL	TCL.cor	RV
30	1.0116	0.9558	1.0449	0.9610	1.0502	1.0379	0.0663	1.0526	1.0484	0.6558	0.6684	1.0524
50	1.0211	0.7053	1.0515	0.9505	1.0494	1.0147	0.2263	1.0495	1.0411	0.7421	0.7232	1.0526
100	0.9926	0.6832	1.0505	1.0105	1.0411	1.0326	0.4779	1.0168	1.0147	0.8179	0.8063	1.0526
200	0.9874	0.6463	1.0516	0.9705	1.0463	1.0358	0.4832	1.0000	0.9989	0.9674	0.8832	1.0516
500	0.9968	0.6558	1.0526	1.0011	1.0421	1.0379	0.5695	0.9989	1.0032	0.9589	0.9737	1.0516

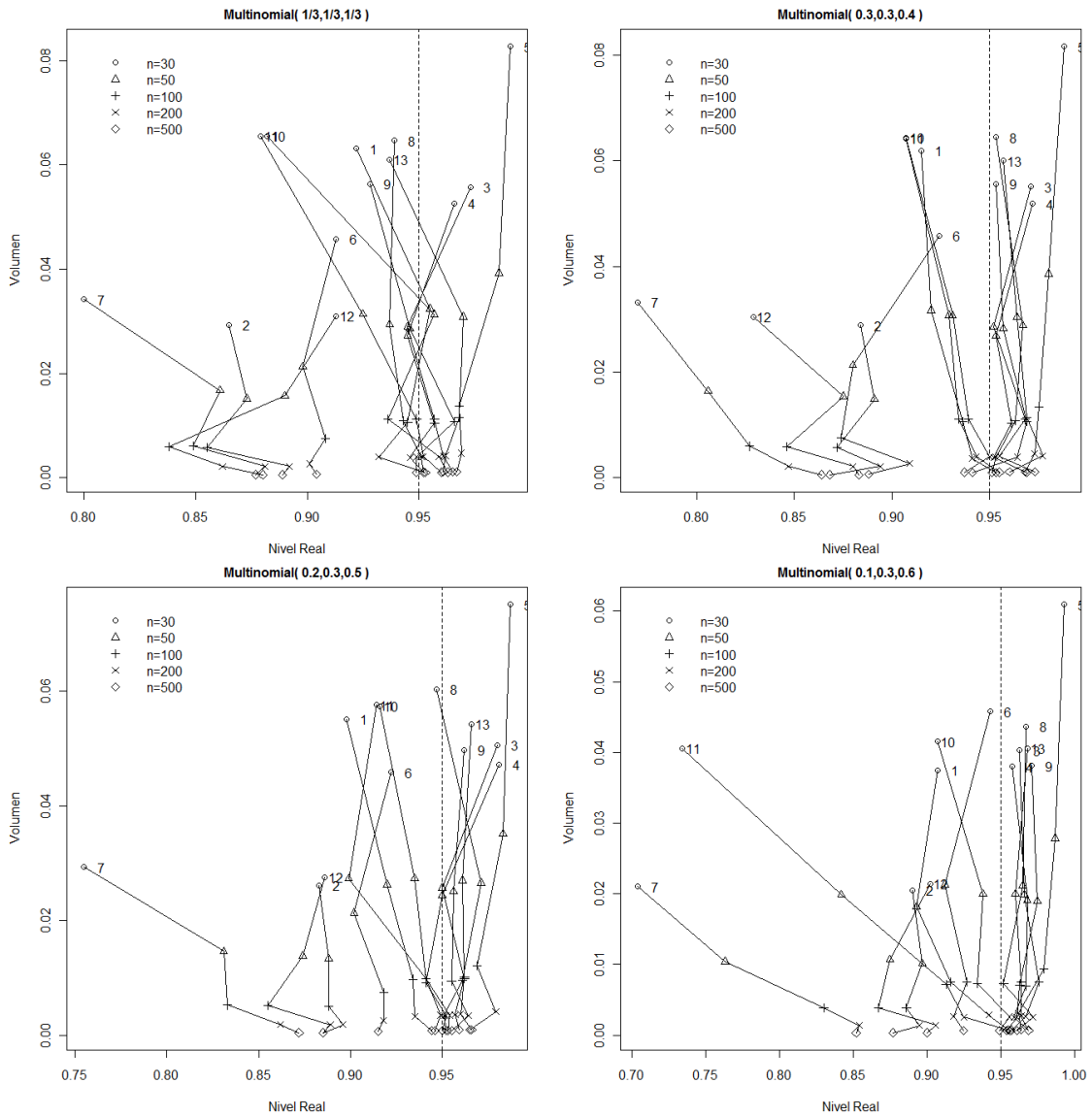


FIGURA 3.1: Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 3

Los métodos aparecen numerados de 1 a 13 en las gráficas <sup>1</sup>

<sup>1</sup>1. IC.SG, 2. IC.Bayes, 3.IC.QH, 4. IC.Goodman, 5. IC.F, 6.IC.Fitz, 7. IC.Boot, 8. IC.Bailey2, 9. IC.Bailey3, 10. IC.TLC, 11. IC.TLC.cor, 12.IC. TLC.II, 13.IC. RV



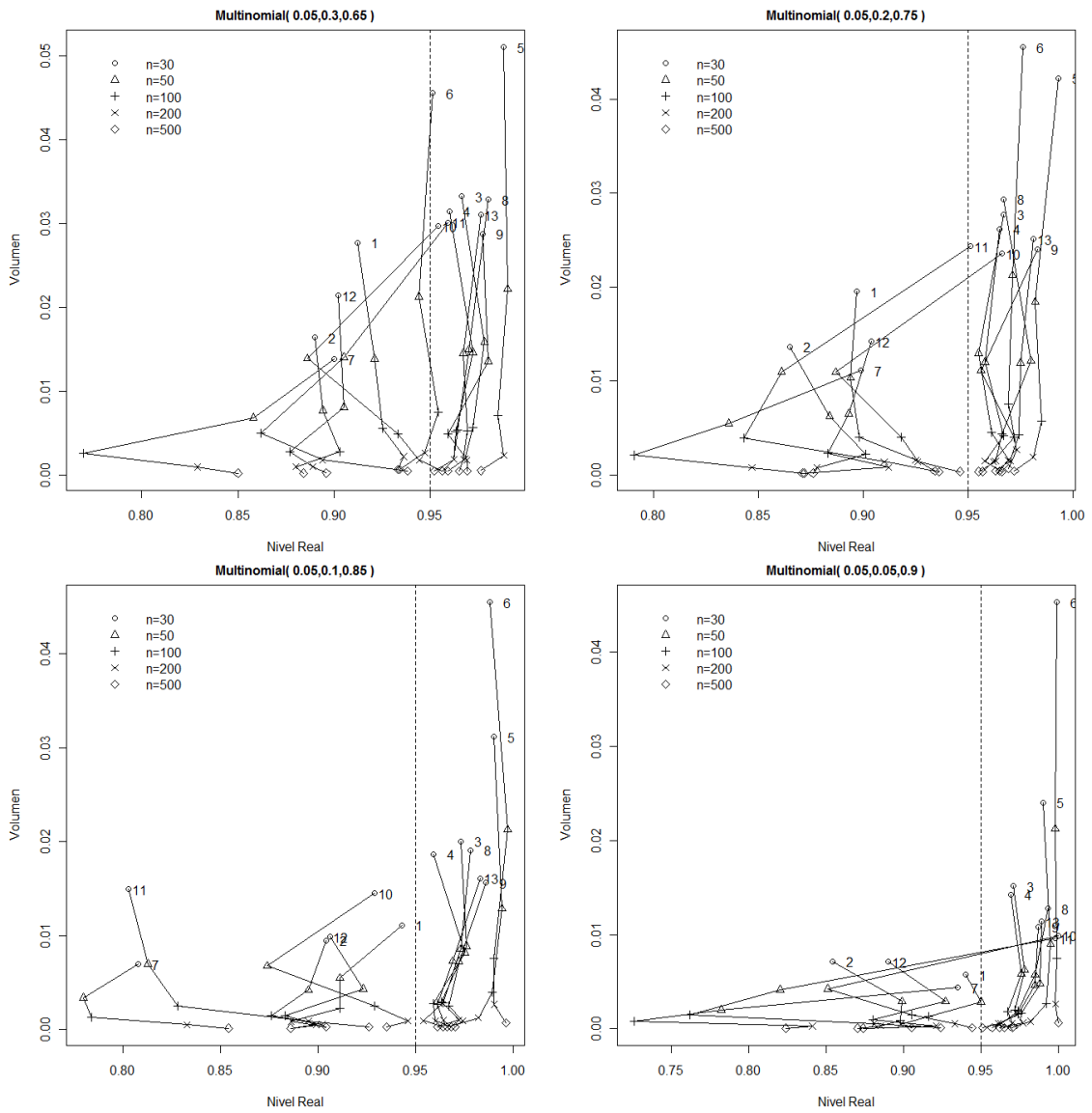


FIGURA 3.2: Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 3

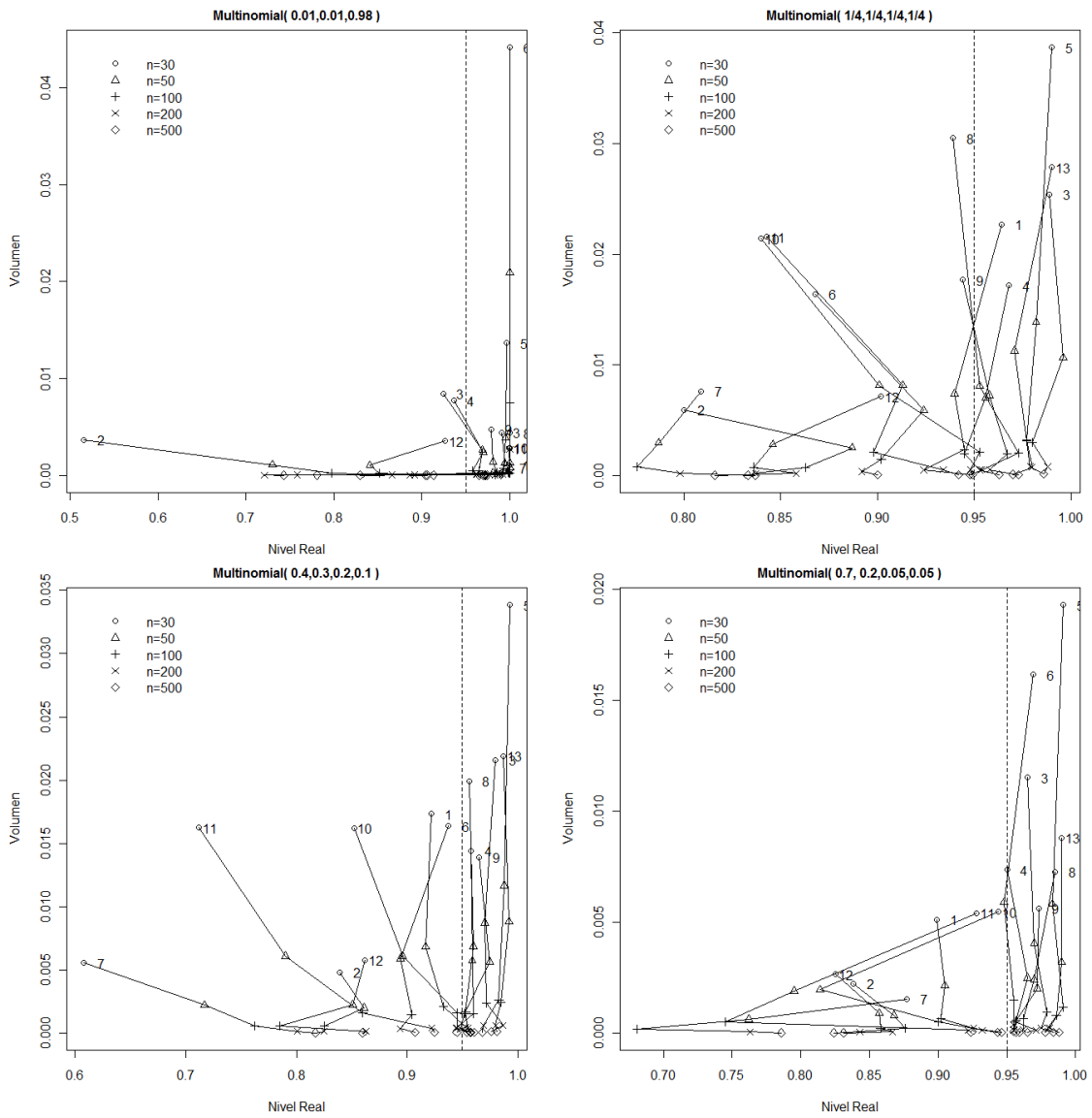


FIGURA 3.3: Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 4

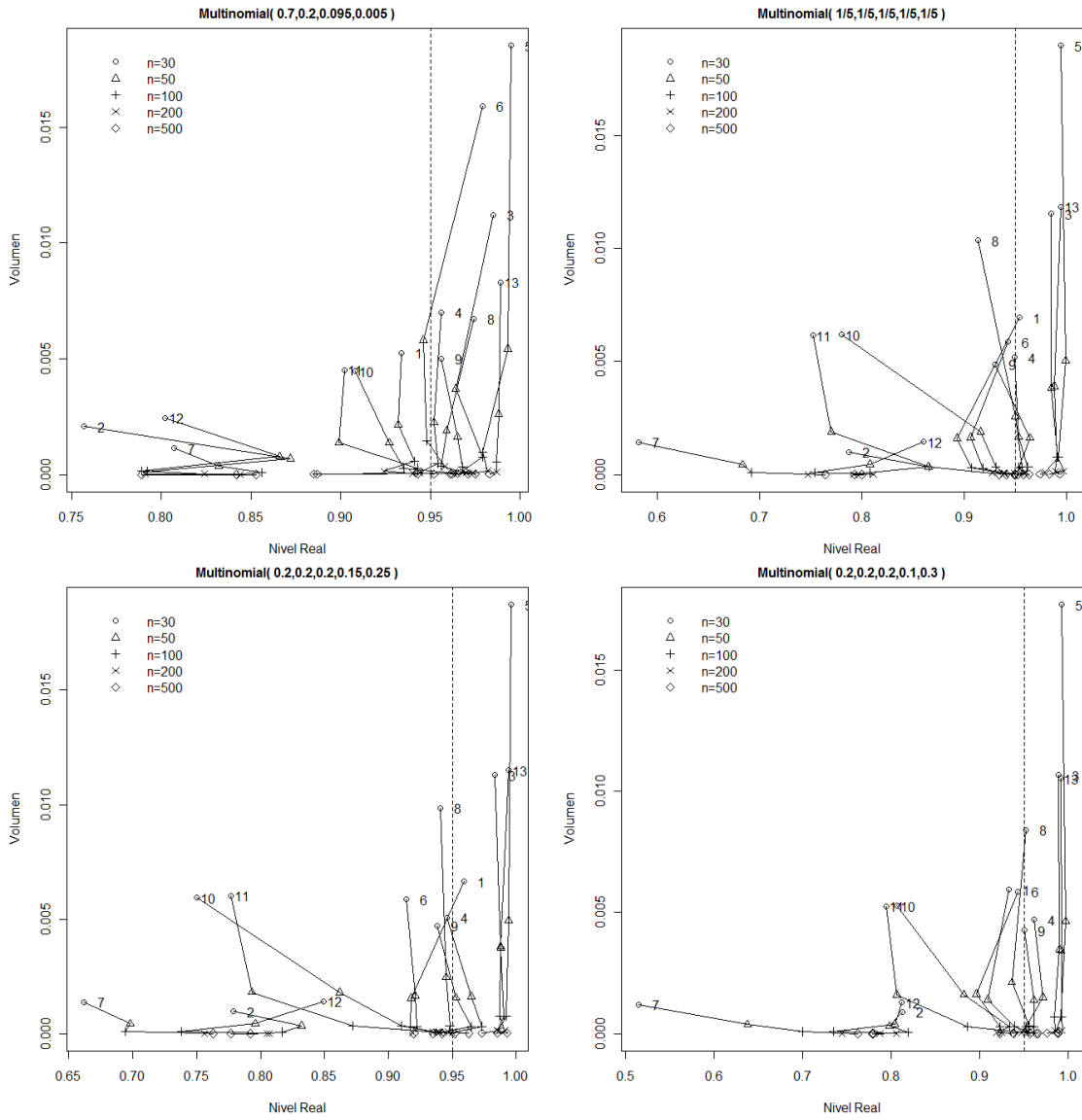


FIGURA 3.4: Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 4 y 5

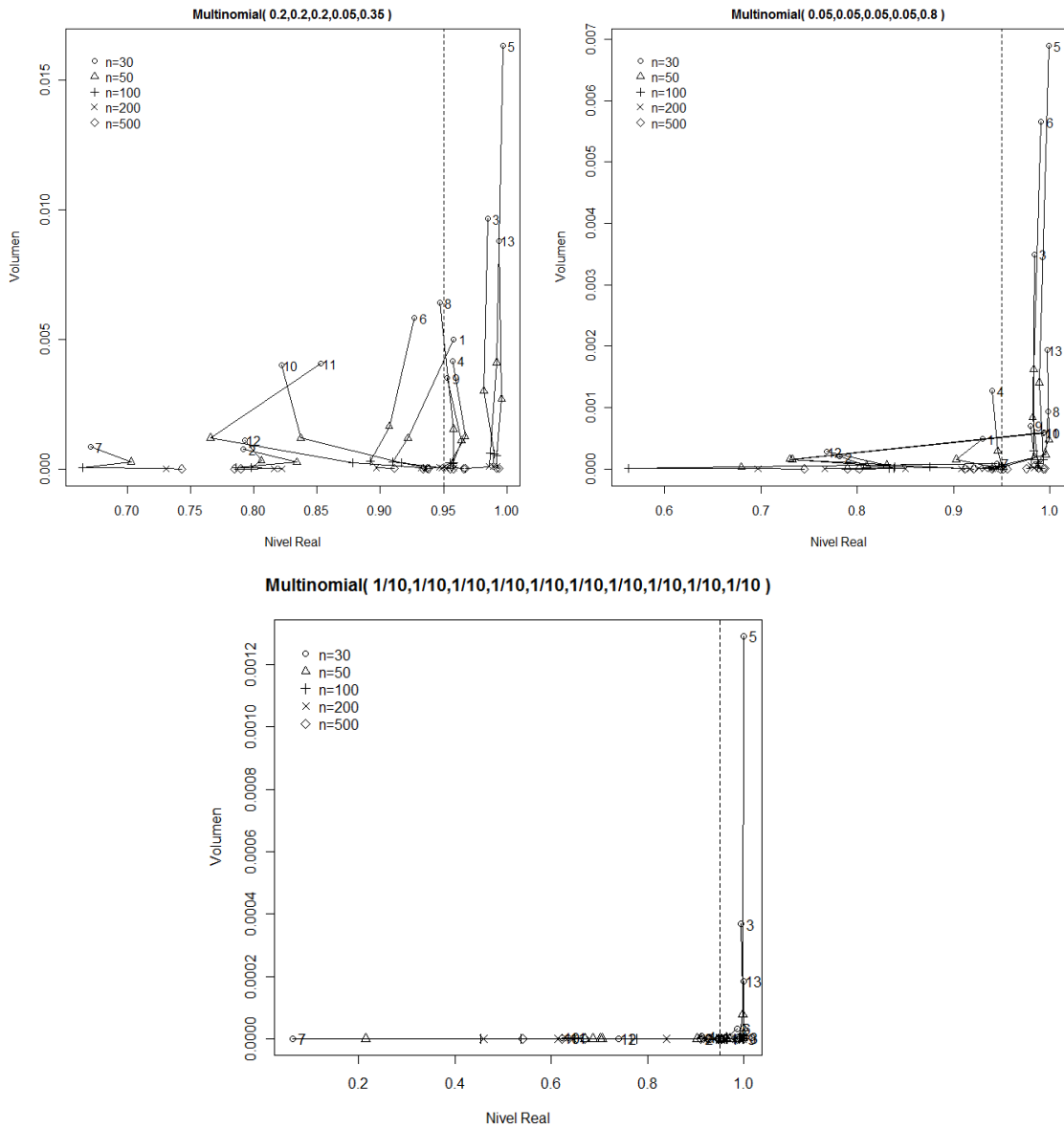


FIGURA 3.5: Resultados gráficos de la simulación para encontrar intervalos de confianza basados en la Distribución Multinomial de dimensión 5 y 10

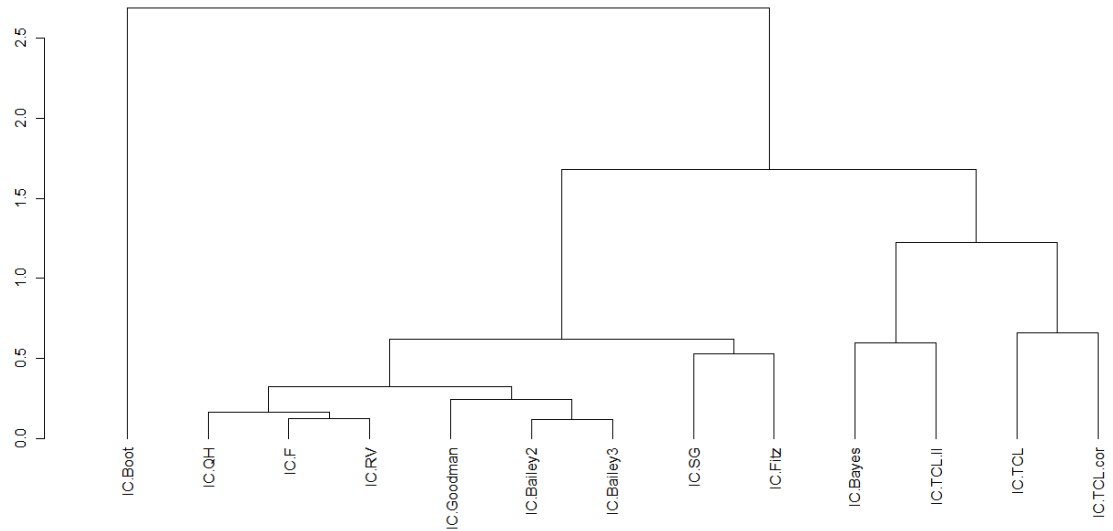


FIGURA 3.6: Cluster con índice

TABLA 3.10: Índice Promedio

Método	Índice Promedio	Orden	Desempeño
<b>F</b>	1.0247	1	Excelente
<b>RV</b>	1.0240	2	
<b>QH</b>	1.0191	3	
<b>Bailey3</b>	1.0071	4	Bueno
<b>Bailey2</b>	1.0048	5	
<b>Goodm</b>	1.0006	6	
<b>Fitz</b>	0.9835	7	
<b>SG</b>	0.9730	8	Regular
<b>TCL</b>	0.9546	9	
<b>TCL.cor</b>	0.9223	10	
<b>TCL.II</b>	0.8833	11	
<b>Bayes</b>	0.8774	12	Deficiente
<b>Boot</b>	0.8012	13	

## 3.2. Resultados y Conclusiones del estudio de simulación

No es recomendable considerar una de las tablas aisladamente para seleccionar el mejor método, ya que no necesariamente un intervalo que tenga un volumen pequeño tiene un nivel de confianza real cercano al nivel nominal, como sucede por ejemplo con los intervalos encontrados con los métodos de Máxima Verosimilitud, método exacto basado en la F y Quesenberry y Hurst, que presentan niveles de confianza reales superiores al nivel nominal pero son los que tienen los volúmenes más grandes del conjunto.

Análogamente un intervalo con un volumen pequeño no necesariamente tiene un nivel de confianza real cercano al nominal como sucede con los intervalos encontrados con los métodos Bootstrap y Bayesiano.

Es importante anotar que el concepto de nivel de confianza real es poco manejado en la práctica, desconociendo su importancia cuando se trabaja con procedimientos que son asintóticos, y también se debe considerar que un nivel de confianza real superior al nivel nominal es mucho mejor que uno por debajo, caso inverso sucede con el volumen del intervalo ya que entre más pequeño mejor es el intervalo. Para trabajar con una combinación de nivel de confianza real y el volumen promedio del intervalo se utiliza el siguiente índice propuesto por Correa & Sierra (2001):

$$I = (1 - Vol) \frac{NR}{NN}$$

donde  $Vol$  es la longitud promedio del intervalo,  $NR$  es el nivel de confianza real y  $NN$  es el nivel de confianza nominal. Este índice es útil para el caso de la distribución multinomial, ya que la longitud de un intervalo estará siempre entre cero y uno. Idealmente la fracción  $NR/NN$  debe estar muy cercana a uno, cabe resaltar que si el volumen del intervalo es grande y la fracción  $NR/NN$  está muy cercana a uno o es mayor que uno el índice no funciona correctamente ya que castigará el método. Por lo tanto entre mayor sea el índice mejor es el método. Las tablas 3.7, 3.8, 3.9 muestran los índices, para cada uno de los trece métodos analizados con diferentes valores de  $N$  y  $\pi$ .

Al analizar todas las tablas y gráficas los intervalos de confianza basados en los métodos de la distribución F, Razón de Verosimilitud, Quesenberry y Hurst alcanzan para todos los tamaños de muestra y cualquier combinación de  $\pi$  niveles de confianza reales que superan el nivel de confianza nominal e

índices muy cercanos a 1. Caso contrario se observa en los intervalos encontrados con el método Bayesiano y Bootstrap que presentan niveles reales por debajo del nivel nominal e índices pequeños.

De las gráficas 3.1 a 3.5 se puede observar que los métodos que aparecen después de la línea vertical punteada, es decir, a la derecha de 0.95 son los que presentan niveles reales cercanos o superiores al nivel nominal, como se puede observar son los métodos basados en la distribución F, Razón de Verosimilitud, Quesenberry y Hurst, Goodman, Bailey2, Bailey3, pero de estos el mejor es el que presenta el volumen más bajo y un índice mayor o igual a 1.

Cabe notar que los intervalos encontrados con el método de Goodman tienen volúmenes más pequeños que los encontrados con el método de Quesenberry, corroborando la conclusión de Goodman (1965).

De las tablas 3.7, 3.8, 3.9 se puede ver que para tamaños de muestra pequeños los intervalos calculados con el método de verosimilitud relativa tienen índices pequeños pero a medida que el tamaño de muestra aumenta el índice crece hasta alcanzar índices mayores o iguales a los encontrados con el método de la F.

Para todos los tamaños de muestra los intervalos encontrados con el método de Quesenberry y Hurst presentan índices buenos y éstos se conservan a medida que el tamaño de muestra aumenta, sin embargo los índices hallados con los dos métodos antes mencionados lo superan.

Del dendograma 3.6 se observa que los intervalos encontrados con el método basado en la F y el método de Razón de Verosimilitud presentan el mínimo nivel de distancia, es decir, la máxima similaridad y cuando se incrementa la distancia los intervalos encontrados mediante el método de Quesenberry y Hurst se unen a este grupo.

Los intervalos calculados con el método de Bootstrap son los más diferentes del grupo ya que presentan la máxima distancia con respecto a los otros métodos.

Si se combina la información de las tablas, gráficos y dendograma observamos que en este último aparecen los métodos clasificados de acuerdo al desempeño, esto se ilustra en la tabla 3.10 clasificando los métodos con desempeño excelente, bueno, regular y deficiente.

## Capítulo 4

# Aplicación: Base de datos de muertes ocurridas en la ciudad de Medellín en 1996

La aplicación de los métodos estudiados se realiza utilizando los datos de los crímenes en la ciudad de Medellín en cada uno de los siete días en una semana seleccionada en forma aleatoria de los meses de Agosto y Septiembre de 1996. Los datos contienen el día, fecha, nombre, sexo, edad, hora y barrio y fueron publicados en la revista “La Hoja” en 1998.

En la tabla 4.1 comparamos los trece intervalos de confianza simultáneos para  $p_i$ ,  $1 \leq i \leq 7$ , con los siguientes tamaños de muestra  $n_1 = 30$ ,  $n_2 = 38$ ,  $n_3 = 44$ ,  $n_4 = 46$ ,  $n_5 = 60$ ,  $n_6 = 80$ ,  $n_7 = 105$  y las proporciones  $p_1 = 0,0744$ ,  $p_2 = 0,0943$ ,  $p_3 = 0,1092$ ,  $p_4 = 0,1141$ ,  $p_5 = 0,1489$ ,  $p_6 = 0,1985$  y  $p_7 = 0,2605$ .

El gráfico 4.1 muestra los intervalos de confianza para cada día de la semana basados en el método exacto de la F, el cual de acuerdo a los resultados de las simulaciones es considerado como el método más adecuado.



TABLA 4.1: Resultados de los intervalos de confianza basados en la Distribución Multinomial aplicados a la base de datos de Muertos en 1996 en la ciudad de Medellín

Método	IC 95 %	Día de la semana							Volumen
		Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	
		$n_1=30$ $p_1=0.0744$	$n_2=38$ $p_2=0.0943$	$n_3=44$ $p_3=0.1092$	$n_4=46$ $p_4=0.1141$	$n_5=60$ $p_5=0.1489$	$n_6=80$ $p_6=0.1985$	$n_7=105$ $p_7=0.2605$	
IC.Bayes	ICI	0.0523	0.0685	0.0809	0.0850	0.1159	0.1624	0.2177	4.7783E-09
	ICS	0.1031	0.1255	0.1410	0.1476	0.1850	0.2384	0.3011	
IC.TCL.II	ICI	0.0526	0.0695	0.08234	0.0866	0.1175	0.1625	0.2201	5.1884E-09
	ICS	0.1043	0.1268	0.14334	0.1489	0.1869	0.2402	0.3055	
IC.Boot	ICI	0.0521	0.0670	0.0794	0.0844	0.1141	0.1588	0.2159	5.3459E-09
	ICS	0.0993	0.1216	0.1414	0.1464	0.1836	0.2457	0.3052	
IC.Bailey3	ICI	0.0437	0.0593	0.0714	0.0755	0.1047	0.1480	0.2038	4.6046E-08
	ICS	0.1140	0.1375	0.1547	0.1605	0.1998	0.2544	0.3209	
IC.Bailey2	ICI	0.0438	0.0595	0.0716	0.0757	0.1051	0.1485	0.2045	4.6721E-08
	ICS	0.1143	0.1378	0.1552	0.1609	0.2003	0.2551	0.3218	
IC.TCL	ICI	0.0393	0.0551	0.0674	0.0715	0.1012	0.1451	0.2017	4.7088E-08
	ICS	0.1096	0.1335	0.1510	0.1568	0.1966	0.2520	0.3194	
IC.TCL.cor	ICI	0.0380	0.0539	0.0661	0.0703	0.0999	0.1438	0.2005	4.7088E-08
	ICS	0.1084	0.1322	0.1497	0.1555	0.1953	0.2507	0.3181	
IC.Goodman	ICI	0.0463	0.0620	0.0741	0.0782	0.1074	0.1506	0.2063	4.8025E-08
	ICS	0.1176	0.1409	0.1581	0.1637	0.2028	0.2571	0.3232	
IC.Fitz	ICI	0.0256	0.0455	0.0604	0.0653	0.1001	0.1497	0.2117	8.4561E-08
	ICS	0.1233	0.1431	0.1580	0.1630	0.1977	0.2473	0.3094	
IC.SG	ICI	0.0248	0.0447	0.0596	0.0645	0.0993	0.1489	0.2109	9.4904E-08
	ICS	0.1241	0.1439	0.1588	0.1638	0.1985	0.2481	0.3102	
IC.F	ICI	0.0431	0.0586	0.0705	0.0745	0.1034	0.1461	0.2013	1.4209E-07
	ICS	0.1297	0.1526	0.1695	0.1751	0.2135	0.2672	0.3326	
IC.RV	ICI	0.0367	0.0510	0.0622	0.0660	0.0935	0.1346	0.1884	3.2036E-07
	ICS	0.1297	0.1543	0.1722	0.1782	0.2188	0.2748	0.3425	
IC.QH	ICI	0.0399	0.0543	0.0655	0.0693	0.0967	0.1376	0.1911	3.3772E-07
	ICS	0.1348	0.1589	0.1766	0.1824	0.2224	0.2777	0.3445	

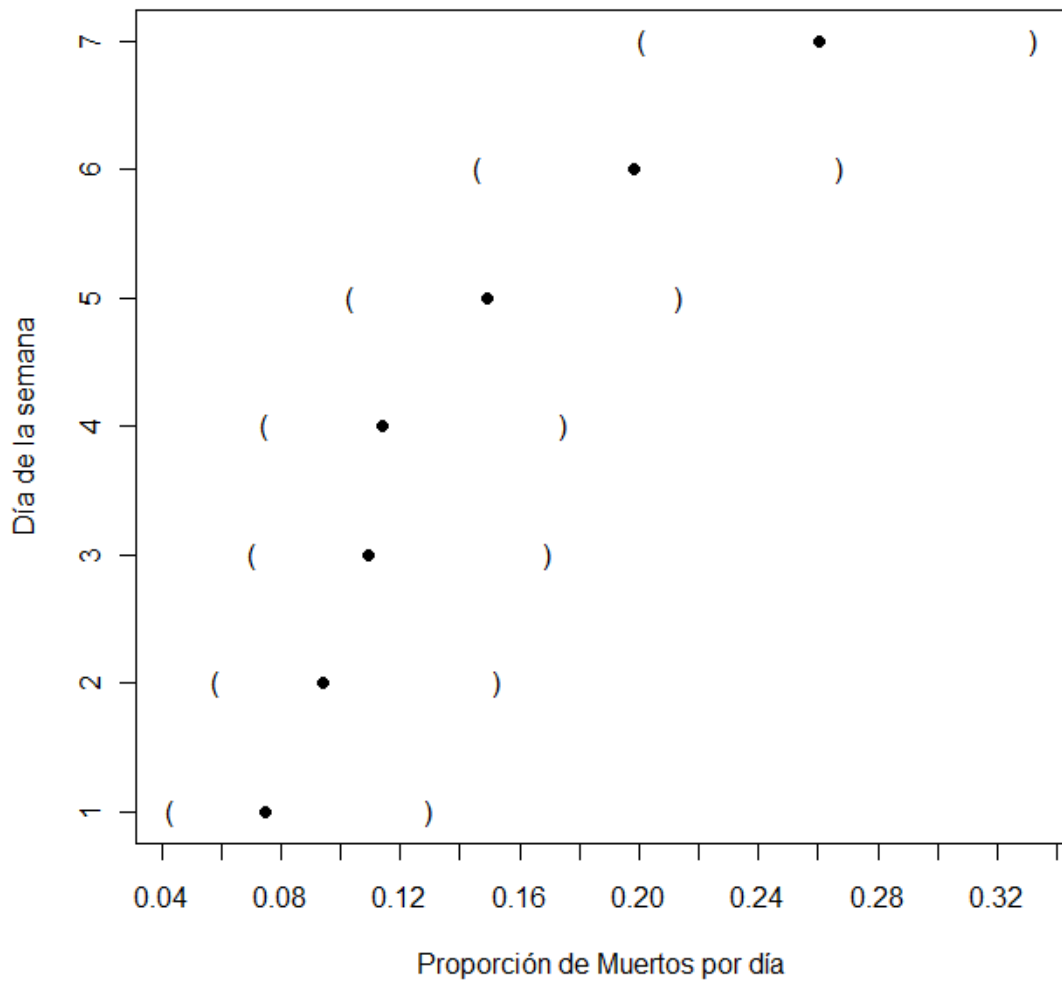


FIGURA 4.1: Intervalos basados en el método exacto de la F

# Capítulo 5

## Conclusiones y Recomendaciones

A lo largo de este trabajo se abordó el problema de los Intervalos de Confianza para la Distribución Multinomial. De los análisis de las simulaciones, es claro que tres procedimientos son superiores; el intervalo de confianza basado en la Función de Verosimilitud, el Intervalo basado en método de la F y el Intervalo basado en el método de Quesenberry y Hurst. El primero de ellos exige encontrar raíces numéricamente, el otro basado en la distribución F es directo y el de Quesenberry y Hurst basado en la aproximación de la distribución chi cuadrada. Lo anterior debido a que los índices encontrados con estos métodos son los más altos del conjunto.

Los otros métodos dan intervalos con índices menores, en particular los intervalos encontrados con los métodos de Bootstrap y el intervalo Bayesiano que presentan los índices más bajos.

De los resultados de las simulaciones llaman la atención los intervalos obtenidos con los métodos basados en el Teorema del Límite Central ya que en la mayoría de los libros de métodos estadísticos los presentan como única alternativa y de acuerdo al análisis que se ha hecho no son los más adecuados.

Este resultado se convierte en un hallazgo importante y da la posibilidad de abrir espacios a nuevas investigaciones y críticas.

# Bibliografía

- Agresti, A. & Caffo, B. (2000), ‘Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures’, *The American Statistician* **54**(4), 280–288.
- Angers, C. (1974), ‘A Graphical Method to Evaluate Sample Sizes for the Multinomial Distribution ’, *Technometrics* **16**(3), 469–471.
- Angers, C. (1979), ‘Sample Size Estimation for Multinomial Populations ’, *The American Statistician* **33**, 163–164.
- Angers, C. (1984), ‘Large Sample Sizes the Estimation of Multinomial Frequencies from Simulation Studies ’, *Simulation* pp. 175–178.
- Angers, C. (1989), ‘Note on Quick Simultaneous Confidence Intervals for Multinomial Proportions ’, *The American Statistician* **43**, 91.
- Bailey, B. (1980), ‘Large Sample Simultaneous Confidence Intervals for the Multinomial Probabilities Based on Transformations of Cell Frequencies ’, *Technometrics* **22**(4), 583–589.
- Bernardo, J. & Smith, A. (2000), *Bayesian Theory*, John Wiley and Sons, Ltd, England.
- Bromaghin, J. (1993), ‘Sample Size Determination for Interval Estimation of Multinomial Probabilities’, *The American Statistician* **47**(3), 203–206.
- Burrows, B. L. (1989), ‘Measures of information and uncertainty ’, *International Journal of Mathematical Education in Science and Technology* **20**, 913–921.
- Canavos, G. (1988), *Probabilidad y Estadística: Aplicaciones y Métodos*, McGraw Hill, Madrid.

- 
- Casella, G. & Berger, R. (2002), *Statistical Inference*, 2 edn, Duxbury, United States of America.
- Cepeda, E., Aguilar, W., Cervantes, V., Corrales, M., Díaz, I. & Rofriguez, D. (2008), 'Intervalos de confianza e intervalos de credibilidad para una proporción', *Revista Colombiana de Estadística* **31**(2), 211–228.
- Chen, H. (1990), 'The Accuracy of Approximate Intervals for a Binomial Parameter', *Journal of the American Statistical Association* **85**(410), 514–518.
- Correa, J. & Sierra, E. (2001), 'Intervalos de confianza para el parámetro de la Distribución Binomial', *Revista Colombiana de Estadística* **24**(1), 59–72.
- Correa, J. & Sierra, E. (2003), 'Intervalos de confianza para la comparación de dos proporciones', *Revista Colombiana de Estadística* **26**(1), 61–75.
- Fitzpatrick, S. & Scott, A. (1987), 'Simultaneous Confidence Intervals for Multinomial Proportions', *Journal of the American Statistical Association* **82**(399), 875–878.
- Glaz, J. & Johnson, B. (1984), 'Probability for Multivariate Distribution with Dependence Structures', *Journal of the American Statistical Association* **79**, 411–436.
- Goodman, L. A. (1965), 'On Simultaneous Confidence Intervals for Multinomial Proportions', *Technometrics* **7**(2), 247–254.
- Hurtubise, R. (1969), 'Sample Sizes and Confidence Intervals Associated with a Monte Carlo Simulation Model Possessing a Multinomial Output', *The American Statistician* **12**, 71–77.
- Kalbfleish, J. (1985), *Probability and Statistical Inference*, 2 edn, Springer-Verlag, New York.
- Leemis, L. & Trivedi, K. (1996), 'A comparison of Approximate Interval Estimators for the Binomial Parameter', *The American Statistician* **50**(1), 63–68.
- Levin, B. (1981), 'A Representation for Multinomial Cumulative Distribution Functions', *The Annals of Statistics* **9**(429), 1123–1126.

- 
- May, W. & Johnson, W. (2000), ‘Constructing Two-Sided Simultaneous Intervals for Multinomial Proportions for Small Counts in a Large Number of Cells’, *Journal of Statistical Software* **5**, 1–24.
- Meyer, P. (1986), *Probabilidad y Aplicaciones Estadísticas*, Addison-Wesley Iberoamericana, México.
- Patel, J. & Read, C. (1996), *Handbook of the Normal Distribution*, Marcel Dekker, Inc, New York.
- Quesenberry, C. & Hurst, D. (1964), ‘Large Sample Simultaneous Confidence Intervals for Multinomial Proportions ’, *Technometrics* **6**(2), 191–195.
- R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Roussas, G. (1973), *A firts Course in Mathematical Statistics*, 2 edn, Addison-Wesley, Massachusetts.
- Sison, C. & Glaz, J. (1995), ‘Simultaneous Confidence Intervals and sample Size determination for Multinomial Proportions’, *Journal of the American Statistical Association* **90**(429), 366–369.
- Snedecor, G. & Cochran, W. (1980), *Statistical Methods*, 7 edn, The Iowa State University Press: Ames, Iowa.
- Thompson, S. (1987), ‘Sample Size for Estimating Multinomial Proportions’, *The American Statistician* **41**, 42–46.
- Tortora, R. (1978), ‘A Note on Sample Size Estimation for Multinomial Populations ’, *The American Statistician* **32**(3), 100–102.
- Walpole, R.E y Myers, R. (1992), *Probabilidad y Estadística*, 4 edn, McGraw Hill, México.

# Apéndice A

## Programas en R

```
#.....  
# PROGRAMAS PARA CALCULAR LOS INTERVALOS DE CONFIANZA DE UNA DISTRIBUCIÓN  
# MULTINOMIAL  
#.....  
  
#.....  
# Cálculo de Intervalos de confianza Basados en Sison y Glaz  
#.....  
  
# Función momentos: Retorna los primeros cuatro momentos y la densidad para un  
# vector de lambdas  
# lambda = np, lambda es un vector de parámetros de la Poisson  
# c1: Es c en el artículo de S&G  
  
momentos<-function(lambda,c1)  
{  
  k=length(lambda)  
  a=lambda+c1  
  b=lambda-c1  
  if( b<0 ) b=0  
  poislama=ppois(a,lambda)  
  poislamb=ppois(b-1,lambda)  
  if( b>0 ) den=poislama-poislamb  
  if( b=0 ) den=poislama  
  .....  
  # Inicialización de matriz con 4 filas para los 4 primeros momentos  
  # Inicialización de matriz con 5 filas para los 4 primeros momentos, la quinta  
  # fila contiene el valor de la densidad  
  .....  
  mu=matrix(0,nrow=4,ncol=1)  
  mom=matrix(0,nrow=5,ncol=1)  
  # r denota el momento.  r=1, media; r=2, varianza; r=3, sesgo; r=4 kurtosis  
  for(r in 1:4)  
  {  
    poisA=0  
    poisB=0  
    if( a-r>=0 ) poisA=poislama-ppois(a-r,lambda)  
    if( a-r< 0 ) poisA=poislama  
    if( b-r-1>=0 ) poisB=poislamb-ppois(b-r-1,lambda)  
    if( b-r-1<0 & b-1>=0 ) poisB=poislamb
```

```

        if( b-r-1<0 & b-1<0 ) poisB=0
        mu[r]=(lambda^r)*(1-(poisA-poisB)/den)
    }
    mu1=mu[1] # 1er momento: media
    mu1_2=mu1^2 # Media al cuadrado
    mu1_3=mu1^3 # Media al cubo
    mu1_4=mu1^4 # Media a la 4
    mu2=mu[2] # 2do momento: varianza
    mu3=mu[3] # 3er momento: sesgo
    mu4=mu[4] # 4to momento: kurtosis
    mom[1]=mu1 # 1er momento: media
    mom[2]=mu2+mu1-mu1_2
    mom[3]=mu3+mu2*(3-3*mu1)+(mu1-3*mu1_2+2*mu1_3)
    mom[4]=mu4+mu3*(6-4*mu1)+mu2*(7-12*mu1+6*mu1_2)+mu1-4*mu1_2+6*mu1_3-3*mu1_4
    mom[5]=den
    return(mom)
} # Fin momentos
.....
# Función truncpoi: Retorna la probabilidad de un vector de poisson truncadas
# x: Es un vector de frecuencias
# c1: Es c en el artículo de S&G
.....
truncpoi<-function(x,c1)
{
    k=length(x)
    n=sum(x)
    m=matrix(0,nrow=k,ncol=5)
    for( i in 1:k)
    {
        lambda=x[i]
        mom<- momentos(lambda,c1)
        for(j in 1:5)
        {
            m[i,j]=mom[j]
        }
    }
    s1=sum(m[,1])
    s2=sum(m[,2])
    s3=sum(m[,3])
    for( i in 1:k)
    {
        m[i,4]=m[i,4]-3*m[i,2]^2
    }
    s4=sum(m[,4])
    probn=1/(ppois(n,n)-ppois(n-1,n))
    z=(n-s1)/sqrt(s2)
    g1=s3/(s2^(3/2))
    g2=s4/(s2^2)
    poly1=1+(g1/6)*((z^3)-3*z)+(g2/24)*((z^4)-6*(z^2)+3)+((g1^2)/72)*((z^6)-15*(z^4)+45*(z^2)-15)
    f=(exp(-(z^2)/2)/sqrt(2*pi))*poly1
    probx=1
    for( i in 1:k)
    {
        probx=probx*m[i,5]
    }
    p=probn*probx*f/sqrt(s2)
    return(p)
} # Fin truncpoi
.....
# Función SG: Retorna los intervalos de S&G

```



```

# n.s: Es un vector de frecuencias
# pis: vector de pis
.....
IC.SG<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  alpha=0.05
  N=sum(n.s)
  k=length(n.s)
  pold=0 # v(c)
  for(c1 in 1:N) # Encontrando el c óptimo
  {
    p<-truncpoi(n.s,c1) # v(c+1)
    if( p <= (1-alpha) | pold >= (1-alpha))
    {
      pold=p
      c11<-c1
    }
    else c1<-N+1
  }
  g=(1-alpha-pold)/(p-pold)
  out=matrix(0,nrow=k,ncol=5)
  num=matrix(0,nrow=k,ncol=1)
  c1=c11
  vol1=1
  vol2=1
  for(i in 1:k)
  {
    num[i,1]=i
    pi.gorro=n.s[i]/N
    cn=c1/N
    onen=1/N
    out[i,1]=pi.gorro
    out[i,2]=pi.gorro-cn
    out[i,3]=pi.gorro+cn+2*g/N
    if( out[i,2]<0) out[i,2]=0
    if( out[i,3]>1) out[i,3]=1
    out[i,4]=pi.gorro-(cn+onen)
    out[i,5]=pi.gorro+(cn+onen)
    if( out[i,4]<0) out[i,4]=0
    if( out[i,5]>1) out[i,5]=1
    IC<- rbind(out[,2],out[,3])
    IC.cor<- rbind(out[,4],out[,5])
  }
  temp1<-NA
  temp2<-NA
  temp3<-NA
  temp4<-NA
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
    temp3[i]<-ifelse(pis[i]>IC.cor[1,i] & pis[i]<IC.cor[2,i],1,0)
    temp4[i]<-IC.cor[2,i]-IC.cor[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0) #
  vol<-prod(temp2)
  level.cor<-ifelse(sum(temp3)==k,1,0)
  vol.cor<-prod(temp4)
}

```

```

    res<-cbind(level.cor,round(vol.cor,30))
    res
} # Fin SG

#.....
# Cálculo de Intervalos de confianza Basados en la aproximación Bayesiana
#.....
# Distribución apriori no informativa  alphas=1
# n: tamaño de muestra
# pis: proporciones (suma de pis debe ser igual a 1)
#.....

IC.Bayes<-function(n.s,pis)
{
  require(coda)
  require(lattice)
  require(MASS)
  library(MCMCpack)
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  k<-length(pis)
  alfa<-n.s+1
  IC<- apply(rdirichlet(10000, alfa),2,quantile,probs=c(0.025,0.975))
  temp1<-NA
  temp2<-NA
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0)
  vol<-prod(temp2)
  res<-cbind(level,round(vol,30))
  res
}

#.....
# Cálculo de Intervalos de confianza de Quesenberry & Hurst
#.....
# n: Frecuencias de las celdas
# pis: proporciones (suma de pis debe ser igual a 1)
#.....

IC.QH<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  alfa<-0.05
  chi<-qchisq(1-alfa,k-1)
  ICI<-(chi+2*n.s-sqrt(chi*(chi+4*(n.s/N)*(N-n.s))))/(2*(N+chi))
  ICS<- (chi+2*n.s+sqrt(chi*(chi+4*(n.s/N)*(N-n.s))))/(2*(N+chi))
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
}

```

```

}
level<-ifelse(sum(temp1)==k,1,0)
vol<-prod(temp2)
res<-cbind(level,round(vol,30))
res
}
#.....
# Cálculo de Intervalos de Confianza de Goodman
#.....

IC.Goodman<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  alfa<-0.05
  chi<-qchisq(1-(alfa/k),1)
  ICI<-((chi+2*n.s-sqrt((chi)*(chi+4*(n.s/N)*(N-n.s)))))/(2*(N+chi))
  ICS<-((chi+2*n.s+sqrt((chi)*(chi+4*(n.s/N)*(N-n.s)))))/(2*(N+chi))
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0) # Para saber si todas las pis caen en los IC
  vol<-prod(temp2) # Para saber el volumen de los IC
  res<-cbind(level,round(vol,30))
  res
}
#.....
# Cálculo de Intervalos de confianza Basados en la Distribución F
#.....

IC.F<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  alfa<-0.05
  k<-length(n.s)
  fi<-qf(1-alfa/(2*k),2*n.s,2*(N-n.s+k-1),lower.tail=F)
  fs<-qf(alfa/(2*k),2*(n.s+k-1),2*(N-n.s),lower.tail=F)
  ICI<- 1/(1+((N-n.s+k-1)/(n.s*fi)))
  ICS<- 1/(1+((N-n.s)/(n.s+k-1)*fs))
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0) # Para saber si todas las pis caen en los IC
  vol<-prod(temp2) # Para saber el volumen de los IC
  res<-cbind(level,round(vol,30))
  res
}

```

```

}
#.....
# Cálculo de Intervalos de Confianza basados en Método de Fitz
#.....

IC.Fitz<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  pi.gorro<-n.s/N
  alfa<-0.05
  ICI<- pi.gorro-qnorm(1-alfa/2)/(2*sqrt(N))
  ICS<- pi.gorro+qnorm(1-alfa/2)/(2*sqrt(N))
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0)
  vol<-prod(temp2)
  res<-cbind(level,round(vol,30))
  res
}
#.....
# Cálculo de Intervalos de Confianza basados en método de Bootstrap
#.....

IC.Boot<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  pi.gorro<-n.s/N
  M=1000
  pi.boot<-rmultinom(M,N,pi.gorro)/N
  ICI<- apply(pi.boot,1,quantile,0.025)
  ICS<- apply(pi.boot,1,quantile,0.975)
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0)
  vol<-prod(temp2)
  res<-cbind(level,round(vol,30))
  res
}
#.....
# Cálculo de Intervalos de Confianza basados en el Método de Bailey2
#.....

```

```

IC.Bailey2<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  alfa<-0.05
  chi<-qchisq(1-(alfa/k),1)
  pib<-(n.s+(3/8))/(N+(3/4))
  ICI<-(sin(asin(sqrt(pib))-sqrt(chi)/sqrt(4*N+2)))^2
  ICS<-(sin(asin(sqrt(pib))+sqrt(chi)/sqrt(4*N+2)))^2
  cota.inf<-(N+3/4)*(sin(chi/sqrt(4*N+2)))^2-3/8
  cota.sup<-(N+3/4)*(sin(pi/2-chi/sqrt(4*N+2)))^2-3/8
  ICI<-ifelse(n.s<=cota.inf,0,ICI)
  ICS<-ifelse(n.s>=cota.sup,1,ICS)
  pib1<-(n.s-0.5+(3/8))/(sum(n.s-.5)+(3/4))
  pib2<-(n.s+0.5+(3/8))/(sum(n.s+.5)+(3/4))
  ICI.cor<-(sin(asin(sqrt(pib1))-sqrt(chi)/sqrt(4*sum(n.s-1/2)+2)))
  ICS.cor<-(sin(asin(sqrt(pib2))+sqrt(chi)/sqrt(4*sum(n.s+1/2)+2)))
  cota.inf.cor<-((sum(n.s-1/2)+3/4)*(sin(chi/sqrt(4*sum(n.s-1/2)+2)))^2)+1/8
  cota.sup.cor<-((sum(n.s+1/2)+3/4)*(sin(pi/2-chi/sqrt(4*sum(n.s+1/2)+2)))^2)-7/8
  ICI.cor<-ifelse(n.s-0.5<=cota.inf.cor,0,ICI.cor)
  ICS.cor<-ifelse(n.s+0.5>=cota.sup.cor,1,ICS.cor)
  IC<- rbind(ICI,ICS,ICI.cor,ICS.cor)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0)
  vol<-prod(temp2)
  res<-cbind(level,round(vol,30))
  res
}
#.....
# Cálculo de Intervalos de Confianza basados en el Método de Bailey3
#.....

IC.Bailey3<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  alfa<-0.05
  Cchi<-qchisq(1-(alfa/k),1)/(4*N)
  pibb<-(n.s+3/8)/(N+1/8)
  ICI<-((sqrt(pibb)-sqrt(Cchi*(Cchi+1-pibb)))^2)/((Cchi+1)^2)
  ICS<-((sqrt(pibb)+sqrt(Cchi*(Cchi+1-pibb)))^2)/(Cchi+1)^2
  cota.inf<-(N+1/8)*Cchi-3/8
  cota.sup<-(N-1/4)
  ICI<-ifelse(n.s<=cota.inf,0,ICI)
  ICS<-ifelse(n.s>=cota.sup,1,ICS)
  Cchi1<-qchisq(1-(alfa/k),1)/(4*sum(n.s-1/2))
  pibb1<-((n.s-1/2)+3/8)/(sum(n.s-1/2)+1/8)
  Cchi2<-qchisq(1-(alfa/k),1)/(4*sum(n.s+1/2))
  pibb2<-((n.s+1/2)+3/8)/(sum(n.s+1/2)+1/8)
}

```

```

ICI.cor<-((sqrt(pibb1)-sqrt(Cchi1*(Cchi1+1-pibb1)))^2)/(Cchi1+1)^2
ICS.cor<-((sqrt(pibb2)+sqrt(Cchi2*(Cchi2+1-pibb2)))^2)/(Cchi2+1)^2
cota.inf.cor<-(sum(n.s-1/2)+1/8)*Cchi1+1/8
cota.sup.cor<-(sum(n.s+1/2)-3/4)
ICI.cor<-ifelse(n.s<=cota.inf.cor,0,ICI.cor)
ICS.cor<-ifelse(n.s>=cota.sup.cor,1,ICS.cor)
IC<- rbind(ICI,ICS,ICI.cor,ICS.cor)
temp1<-NULL
temp2<-NULL
for(i in 1:k)
{
  temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
  temp2[i]<-IC[2,i]-IC[1,i]
}
level<-ifelse(sum(temp1)==k,1,0) # Para saber si todas las pis caen en los IC
vol<-prod(temp2) # Para saber el volumen de los IC
res<-cbind(level,round(vol,30))
res
}

#.....
# Cálculo de Intervalos de Confianza basados en el Teorema Central del Límite
#.....

IC.TCL<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  pi.gorro<-n.s/N
  k<-length(n.s)
  alfa<-0.05
  ICI<- pi.gorro-qnorm(1-alfa/(2*k))*sqrt((pi.gorro*(1-pi.gorro))/N)
  ICS<- pi.gorro+qnorm(1-alfa/(2*k))*sqrt((pi.gorro*(1-pi.gorro))/N)
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0)
  vol<-prod(temp2)
  res<-cbind(level,round(vol,30))
  res
}

#.....
# Cálculo de Intervalos de Confianza basados en el Teorema Central del Límite
# Corregido
#.....
IC.TCL.cor<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  pi.gorro<-n.s/N
  k<-length(n.s)
  alfa<-0.05
  ICI<- pi.gorro-qnorm(1-alfa/(2*k))*sqrt((pi.gorro*(1-pi.gorro))/N)-(1/(2*N))

```

```

ICS<- pi.gorro+qnorm(1-alfa/(2*k))*sqrt((pi.gorro*(1-pi.gorro)/N)-(1/(2*N))
IC<- rbind(ICI,ICS)
temp1<-NULL
temp2<-NULL
for(i in 1:k)
{
  temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
  temp2[i]<-IC[2,i]-IC[1,i]
}
level<-ifelse(sum(temp1)==k,1,0)
vol<-prod(temp2)
res<-cbind(level,round(vol,30))
res
}
#.....
# Cálculo de Intervalos de Confianza basados en el Teorema Central del Límite II
#.....
IC.TCL.II<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  pi.gorro<-n.s/N
  k<-length(n.s)
  alfa<-0.05
  ICI<- pi.gorro*(N/(N+(qnorm(1-alfa/2)^2)))+(1/2)*((qnorm(1-alfa/2)^2)/(N+(qnorm(1-alfa/2)^2)))-
    sqrt((1/(N+(qnorm(1-alfa/2)^2)))*(pi.gorro*(1-pi.gorro)*
      (N/(N+(qnorm(1-alfa/2)^2)))+(1/4)*((qnorm(1-alfa/2)^2))))
  ICS<- pi.gorro*(N/(N+(qnorm(1-alfa/2)^2)))+(1/2)*((qnorm(1-alfa/2)^2)/(N+(qnorm(1-alfa/2)^2)))+
    sqrt((1/(N+(qnorm(1-alfa/2)^2)))*(pi.gorro*(1-pi.gorro)*
      (N/(N+(qnorm(1-alfa/2)^2)))+(1/4)*((qnorm(1-alfa/2)^2))))
  IC<- rbind(ICI,ICS)
  temp1<-NULL
  temp2<-NULL
  for(i in 1:k)
  {
    temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
    temp2[i]<-IC[2,i]-IC[1,i]
  }
  level<-ifelse(sum(temp1)==k,1,0) # Para saber si todas las pis caen en los IC
  vol<-prod(temp2) # Para saber el volumen de los IC
  res<-cbind(level,round(vol,30))
  res
}
#.....
# Cálculo de Intervalos de Confianza basados en el Método de Razón de Verosimilitud
#.....
# Hallar raíz
IC.RV<-function(n.s,pis)
{
  n.s<-ifelse(n.s==0,0.5,n.s)
  n.s<-ifelse(n.s==sum(n.s),n.s-0.5,n.s)
  N<-sum(n.s)
  k<-length(n.s)
  pi.g<-n.s/N
  log.L<-function(pi.v,pi.g,n.s,N) -2*(n.s*log(pi.g/pi.v)+(N-n.s)*
    log((1-pi.g)/(1-pi.v)))+ qchisq(0.95,k-1)

```

```

pi.v<-seq(0.001,0.999,length=100)
res<-NULL
ICI<-NULL
ICS<-NULL
for(i in 1:k)
{
  res<-sapply(pi.v,log.L,pi.g[i],n.s[i],N)
  ICI[i]<-uniroot(log.L,c(0,pi.g[i]),pi.g[i],n.s[i],N)$root
  ICS[i]<-uniroot(log.L,c(pi.g[i],1),pi.g[i],n.s[i],N)$root
}
IC<- rbind(ICI,ICS)
temp1<-NULL
temp2<-NULL
for(i in 1:k)
{
  temp1[i]<-ifelse(pis[i]>IC[1,i] & pis[i]<IC[2,i],1,0)
  temp2[i]<-IC[2,i]-IC[1,i]
}
level<-ifelse(sum(temp1)==k,1,0)
vol<-prod(temp2)
res<-cbind(level,round(vol,30))
res
}

#.....
# Código para las Simulaciones
#.....
require(lattice)
require(coda)
require(MASS)
library(MCMCpack)

pis<-matrix(c(1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10) ,ncol=10,byrow=T)
#(1/5,1/5,1/5,1/5,1/5,
#0.2,0.2,0.2,0.15,0.25,
#0.2,0.2,0.2,0.1,0.3,
#0.05,0.05,0.05,0.05,0.8,
#0.2,0.2,0.2,0.05,0.35) ,ncol=5,byrow=T)

#(1/4,1/4,1/4,1/4,
#0.4,0.3,0.2,0.1,
#0.7, 0.2,0.05,0.05,
#0.7,0.2,0.095,0.005) ,ncol=4,byrow=T)
#(1/3,1/3,1/3,
#.3,.3,.4, #) ,ncol=3,byrow=T)
#.2,.3,.5, #!!!!
#.1,.3,.6, #!!!!
#.05,.3,.65,
#.05,.2,.75,
#.05,.1,.85,
#.05,.05,.9,
#.01,.01,.98) ,ncol=3,byrow=T)
funcion<-c('IC.SG','IC.Bayes','IC.QH','IC.Goodman','IC.F','IC.Fitz',
'IC.Boot','IC.Bailey2','IC.Bailey3','IC.TCL','IC.TCL.cor',
'IC.TCL.II','IC.RV')

aplique<-function(pis,funcion)
{
  N<-matrix(c(30, 50, 100, 200, 500),ncol=1)
  Nsim=1000

```



```

temp<-NULL
aux<-NULL
res<-NULL
for(i in 1:nrow(N))
{
  temp<-rmultinom(Nsim,N[i,],pis)
  res<- t(rowMeans(apply(temp,2,funcion,pis)))
  aux<-cbind(aux,res)
}
aux
}
res1<-NULL
res4<-NULL
for(i in 1:nrow(pis))
{
  for(j in 1:length(funcion))
  {
    res1<-aplique(pis[i,],funcion[j])
    res4<-rbind(res4,res1)
  }
}
write.table(res4,'D:/Documents/Multinomialfin/resfinal.csv',sep=',',dec='.')

#.....
# Código para las Cluster y Gráficos
#.....
datos<-read.csv2('D:/Documents/Multinomialfin/formatocluster.csv',sep=',',dec='.')
attach(datos)
dim(datos)
rownames(datos)<-datos[,1]
hc <- hclust(dist(datos[,-1]))
plot(hc)
(dend1 <- as.dendrogram(hc))
plot(dend1)
#Gráficos
setwd('D:/Documents/Multinomialfin')
datos<-read.csv2("formato para analisis.csv",sep=',',dec='.')
attach (datos)
str(datos)
grafica1<-function(datos,titulo)
{
a.vol<-min(datos[,5])
b.vol<-max(datos[,5])
a.lev<-min(datos[,4])
b.lev<-max(datos[,4])
for(i in 1:13)
{
  pch.i<-1:5
  aux<-datos[metodo==i,]
  if(i==1)
  plot(aux[,4],aux[,5],type='l',ylim=c(a.vol,b.vol),xlim=c(a.lev,b.lev),ylab='Volumen',
  xlab='Nivel Real',main=(paste("Multinomial(",titulo,")"))
  else
  lines(aux[,4],aux[,5])
  points(aux[,4],aux[,5],pch=pch.i,cex=1)
  legend(aux[aux[,3]==30,4][1],aux[aux[,3]==30,5][1],legend=i,bty='n',
  xjust=.5,yjust=0.5)
}
abline(v=.95,lty=2)

```

```
    legend(a.lev,b.vol,legend=c('n=30','n=50','n=100','n=200','n=500'),
           pch=pch.i,bty='n')
}
par(mfrow=c(1,2),mar=c(4,4,2,2),cex.main=1)
titulo<-c('1/3,1/3,1/3',
'0.3,0.3,0.4',
'0.2,0.3,0.5',
'0.1,0.3,0.6',
'0.05,0.3,0.65',
'0.05,0.2,0.75',
'0.05,0.1,0.85',
'0.05,0.05,0.9',
'0.01,0.01,0.98',
'1/4,1/4,1/4,1/4',
'0.4,0.3,0.2,0.1',
'0.7, 0.2,0.05,0.05',
'0.7,0.2,0.095,0.005',
'1/5,1/5,1/5,1/5,1/5',
'0.2,0.2,0.2,0.15,0.25',
'0.2,0.2,0.2,0.1,0.3',
'0.2,0.2,0.2,0.05,0.35',
'',
'0.05,0.05,0.05,0.05,0.8',
'1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10')
l<-length(unique(datos[,1]))+1
for(j in 1:l)
{
  datos.aux<-datos[is==j,]
  grafical(datos.aux,titulo[j])
}
```