

NOTA TÉCNICA

## COMPAPROP: Sistema para comparación de proporciones múltiples

Yoannia Castillo Duvergel<sup>I</sup>, Ileana Miranda<sup>II</sup>

<sup>I</sup>Dirección de Ciencia, Innovación y Postgrado. <sup>II</sup>Dirección Sanidad Vegetal. Centro Nacional de Sanidad Agropecuaria (CENSA), Apartado 10, San José de las Lajas, Mayabeque. CP: 32700. Cuba. Correo electrónico: yoannia@censa.edu.cu.

**RESUMEN:** Se desarrolló un software, implementado en macros con lenguaje Microsoft Visual Basic para Aplicaciones y Framework Grails, que permite comparar proporciones múltiples por los métodos: Wald, Wald con corrección por continuidad, McNemar y Few Test. El sistema COMPAPROP brinda una forma rápida y eficaz de realizar una selección de tratamientos sobre la base de su contraste.

**Palabras clave:** intervalo de confianza, método Wald, Wald con corrección por continuidad, McNemar, Few Test.

---

### COMPAPROP: A system for multiple proportion comparisons

**ABSTRACT:** A software was developed, implemented in macros using Microsoft Visual Basic for Applications and Grails Framework, allowing to compare multiple proportions by the methods Wald, Wald with continuity correction, McNemar and Few Test. COMPAPROP System provides a fast and efficient way for carrying out a variety of treatments based on its contrast.

**Key words:** confidence intervals, Wald method, Wald with continuity correction, McNemar, Few Test.

---

El uso de tecnologías y procedimientos modernos en la estadística aplicada, que faciliten el trabajo de los usuarios, resulta una necesidad en entidades agropecuarias; por lo que colocar herramientas al servicio de la investigación representa un desafío en la actualidad.

Existen antecedentes de software que incluyen la comparación de proporciones como el «SAS» y el «SPSS», pero estos sólo presentan la posibilidad de contraste de hipótesis para dos proporciones (1), y en el caso del paquete estadístico «R», el software se proporciona con derechos limitados, y algunas de las funciones avanzadas del sistema sólo son accesibles a través de la sintaxis de comandos y no están disponibles en la versión estudiantil, gratuita y de libre acceso (2). Esto limita la obtención de resultados para proporciones múltiples, lo que obliga al usuario a utilizar herramientas en plataformas ya en desuso como MS-DOS. Por ello, se hizo necesario el desarrollo de una herramienta que supere estas deficiencias y brinde la posibilidad de tomar mejores decisiones, cuando

se tiene una selección de tratamientos sobre la base de su contraste.

Para el desarrollo de COMPAPROP («Comparación de proporciones múltiples») se utilizó como base, el cálculo de intervalos de confianza y no las pruebas de hipótesis, los mismos aportan información de la magnitud y la precisión de las estimaciones, y se puede interpretar el intervalo en términos del margen de error de la estimación puntual. Esto hace los intervalos más atractivos a la hora de presentar resultados, mientras que el valor-p en las pruebas de hipótesis, es una elaboración probabilística de interpretación más compleja (3). Además, se hizo un estudio referencial de los algoritmos utilizados en este aspecto y se consideró, para la implementación, los métodos de Wald, Wald con corrección por continuidad, McNemar y Few Test.

Notación: Sean  $x_{i_1}, x_{i_2}, \dots, x_{i_{n_i}}$  para  $i=1,2$ , muestras aleatorias  $i$ , independientes, de tamaño  $n_i$  de distribuciones Bernoulli con parámetros  $\pi_i$ . El estimador de máxima verosimilitud para  $\pi_i$  está dado por:

$$\hat{\Pi}_1 = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \frac{x_i}{n_j} \text{ Con } x_i \text{ el número de éxitos}$$

en la muestra  $i$ .

Método Wald: este método aparece en la mayoría de los textos básicos de Estadística, se basa en la distribución asintótica normal de la diferencia entre las proporciones muestrales,  $\pi_1 - \pi_2$ . Teóricamente, el método brinda un intervalo con un nivel de confianza aproximado de  $(1 - \alpha) \cdot 100\%$ , la aproximación es mejor en tanto  $n_1$  y  $n_2$  sean mayores que 30. (3)

$$\Pi_1 - \Pi_2 \pm z \frac{\alpha}{2} \sqrt{\frac{\Pi_1(1-\Pi_1)}{n_1} + \frac{\Pi_2(1-\Pi_2)}{n_2}}$$

Donde:

$z \frac{\alpha}{2}$  es el percentil  $100 \cdot (1 - \frac{\alpha}{2})$  de la normal estándar.(3)

Método Wald con corrección por continuidad: método eficiente, donde se le hace modificaciones al método Wald. Como la aproximación normal a la binomial es de una distribución discreta por una continua, es pertinente hacer la corrección por continuidad. (3)

Los extremos de este intervalo son:

Límite del intervalo:

$$\Pi_1 - \Pi_2 \pm (z \frac{\alpha}{2} \sqrt{\frac{\Pi_1(1-\Pi_1)}{n_1} + \frac{\Pi_2(1-\Pi_2)}{n_2}} + (\frac{1}{n_1} + \frac{1}{n_2}) / 2$$

Método McNemar: es una variación del Test *Sign* para significación entre proporciones, el cual quizás fue el primer test no-paramétrico publicado por Arbuthnott, 1710 (4).

Límite del intervalo:

$$\frac{1}{2^n} \cdot \sum_{i=0}^{\alpha} C_i^n < \frac{\alpha}{2} \text{ ó } \frac{1}{2^n} \cdot \sum_{i=0}^{\alpha} C_i^n > 1 - \frac{\alpha}{2}$$

Donde:

$$C_i^n = \frac{n!}{i! \cdot (n-i)!}$$

Método Few Test: es un nuevo Test estadístico para la comparación de proporciones, desarrollado por

University of Applied Sciences of Western Switzerland (4).

$$\hat{T} = \max(\sum_{i=a}^{n_a} \sum_{j=0}^b C_i^{n_a} \cdot C_j^{n_b} \cdot p^{i+j} \cdot (1-p)^{n_a+n_b-i-j}) + (\sum_{i=0}^{n_a} \sum_{j=n_b-b}^{n_b} C_i^{n_a} \cdot C_j^{n_b} \cdot p^{i+j} \cdot (1-p)^{n_a+n_b-i-j}) < \alpha$$

Para validación de intervalos de confianza se consideró la precisión, indicada por la longitud del intervalo y la probabilidad de cobertura:

$$P(L_{inf} \leq \Pi_1 - \Pi_2 \leq L_{sup})$$

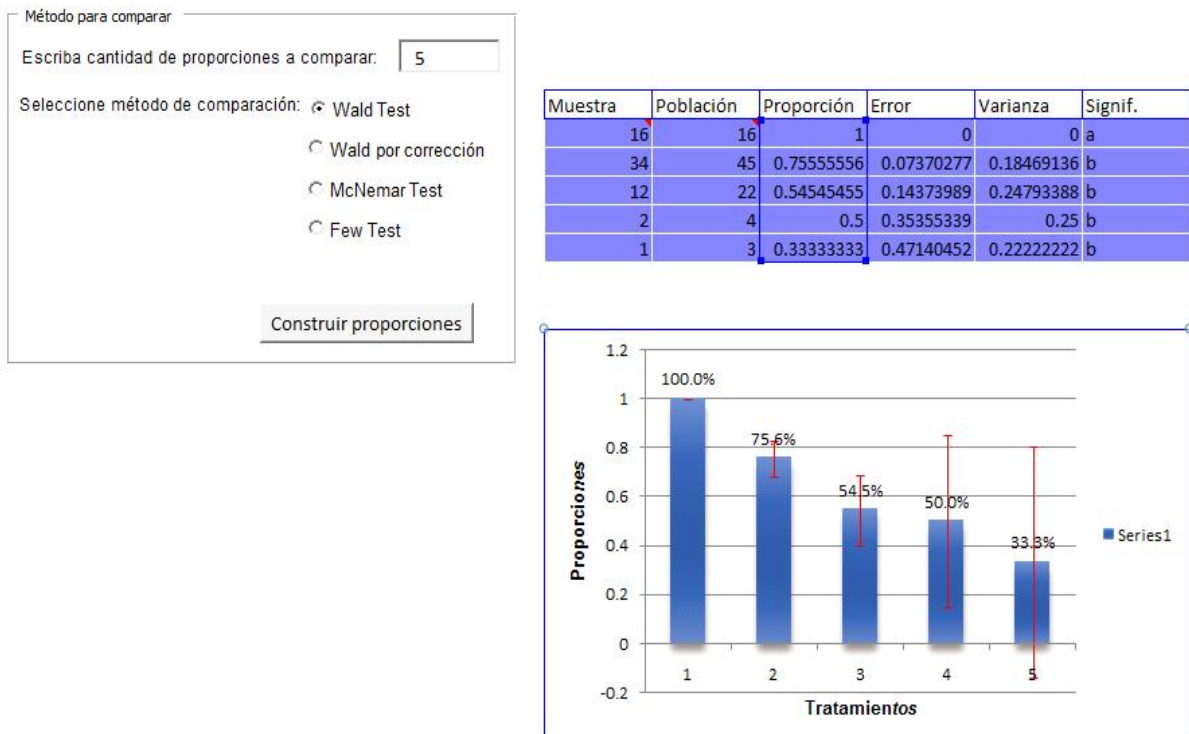
Se buscó que los algoritmos usados para construir los intervalos de confianza ofrecieran intervalos tales que:

1. Sus longitudes fueran pequeñas, pero diferentes de cero.
2. La probabilidad de cobertura no fuera inferior al nivel de confianza nominal.

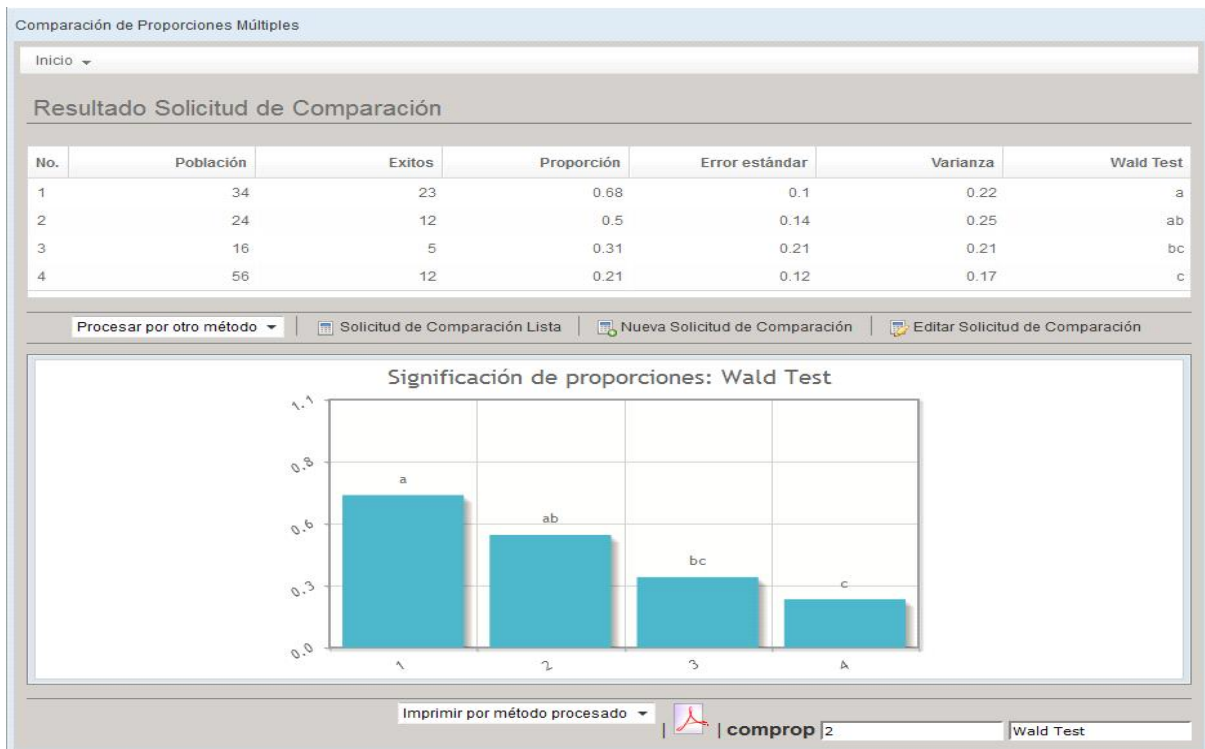
El software se implementó en dos versiones: utilizando macros con lenguaje *Visual Basic for Application* 6.0 (VBA): control remoto para Microsoft Excel (5). Con esta versión se hizo factible la integración con Microsoft Excel 2007, y dio la posibilidad al usuario de seguir realizando otros cálculos y gráficos (Fig. 1).

La segunda versión se implementó con Framework Grails: marco de trabajo para el desarrollo de aplicaciones web que se basa en las ideas de «Codificación a través de Convenciones» (*Coding by Convention*) y DRY (*Don't Repeat Yourself*) utilizando el lenguaje de programación Groovy (6,7). Se utilizó, además, para las interfaces web de usuario, ZK: *framework* de aplicaciones web en AJAX (8). Esta versión brinda la posibilidad de crear sus propias solicitudes de comparación para el posterior análisis, el acceso y la utilización vía web sin necesidad de instalarlo en la máquina (Fig. 2).

A pesar de los avances teóricos y computacionales en la estadística de las últimas décadas, se prestó poca atención a uno de los problemas más comunes en la investigación estadística: el análisis de la diferencia de proporciones. Lo anterior implica que, para este problema específico, se planteó esta solución computacional, la cual se encuentra disponible para especialistas de Sanidad Agropecuaria, así como para estudiantes universitarios y otros interesados.



**FIGURA 1.** Versión1. Comparación de proporciones múltiples en macros./ *Version 1. Multiple proportion comparison in macros.*



**FIGURA 2.** Versión2. Comparación de proporciones múltiples en Framework Grails./ *Version 2. Multiple proportion comparison in Grails framework.*

**REFERENCIAS**

1. Rodríguez MS, Montes I. Curso Avanzado del Paquete Estadístico R. Introducción a la modelización estadística. Universidad de Oviedo. E.U. Jovellanos - Campus de Viesques. 2009.
2. SPSS Inc. SPSS Statistics Base User's Guide 19.0. Copyright © 2010.
3. Correa JC, Sierra E. Intervalos de confianza para la comparación de dos proporciones. Revista Colombiana de Estadística. 2003;26(1).
4. Taillard DE, Waelti P, Zuber J. Few statistical tests for proportions comparison. University of Applied Sciences of Western Switzerland. HEIG-VD campus at Yverdon Route de Cheseaux 1, case postale CH-1401 Yverdon-les-Bains, Switzerland. 2008.
5. Walkenbach J. Excel® 2007 Bible. Copyright © by Wiley Publishing, Inc., Indianapolis, Indiana. 2007.
6. Smith G, Ledbrook P. Grails in Action. Copyright © by Manning Publications Co. All rights reserved. 2009.
7. König D, Glover A, King P, Laforge G, Skeet J. Groovy in Action. Copyright © by Manning Publications Co. All rights reserved. 2007.
8. ZK Developer's Reference. [Citado 13 de septiembre 2012]. Disponible en: <http://code.pediapress.com>.

Recibido: 10-6-2014.

Aceptado: 15-11-2014.