

LA APLICACIÓN DE MÉTODOS ESTADÍSTICOS EN LAS CIENCIAS HUMANÍSTICAS Y SU ENSEÑANZA

MSc. Teresa Pérez Sosa¹, Dr. Roberto Suarez Ojeda², MSc. Adriana Delgado Landa⁴, Lic. Dirma Yanes Quintero⁴.

*1.2.3.4. Universidad de Matanzas “Camilo Cienfuegos”, Vía Blanca
Km.3, Matanzas, Cuba.*

RESUMEN

En las ciencias sociales durante los últimos años ha tenido avances en aplicaciones estadísticas, lo que favorece una absorción de poca calidad por parte de los especialistas en áreas no metodológicas.

La necesidad de uso de los recursos estadísticos, consigue finalmente que el especialista en áreas aplicadas tienda a descuidar aspectos muy básicos, previos a la aplicación de estos recursos estadísticos complejos.

Sin embargo, en dichas ciencias es frecuente, más aún, predominante, la existencia de variables que no corresponden a una escala de intervalo o razón, es decir, cuyos valores no constituyen verdaderamente mediciones, siendo lo más frecuente el trabajo con variables correspondientes al nivel de medida nominal u ordinal. En dicha situación no resultan aplicables las técnicas mencionadas anteriormente, sino otras, incluidas dentro de los llamados “métodos no paramétricos” y otras técnicas tales como la regresión ordinal y logística. El trabajo examina esta cuestión, sugiriendo los temas y técnicas que se consideran más apropiados en carreras e investigaciones de este perfil, describiendo algunas de ellas.

Palabra clave: ciencias sociales, métodos, no paramétricos.

INTRODUCCIÓN

Cada vez es mayor la proporción de investigadores, en las más diversas disciplinas científicas, que realizan análisis estadísticos de datos como procedimiento formal para llegar a conclusiones o apoyar procesos de decisión sobre las hipótesis de la investigación. Para muchos de ellos, utilizar las herramientas de la Estadística en la investigación, dentro de su área de conocimiento, es un quehacer poco motivador, puesto que su dominio puede alejarse bastante de los objetivos de formación, investigación y aplicación del experto en esa área. La duda que se genera tanto en el investigador como en el docente es: «Cuando se estudia estadística, no para convertirse en especialista en la materia sino para trabajar en ciencias humanas, en psicología en concreto, ¿hasta dónde habría que profundizar en aquella materia?»

Este dilema entre motivación y necesidad, en la utilización de los recursos estadísticos tiene consecuencias importantes. «La necesidad de acudir a las técnicas estadísticas matemáticas para la resolución científica de la mayor parte de los problemas plantea una dificultad a veces insuperable. ¿Cómo puede un investigador usar dichas técnicas si carece de conocimientos sólidos de la ciencia matemática?».

Los recursos de la estadística para la investigación científica pueden considerarse como un conjunto de herramientas conceptuales que poseen sus propias condiciones de aplicación y pautas de uso. Respetando ambos aspectos, las conclusiones que se realizan sobre las informaciones tratadas convenientemente con herramientas estadísticas, tendrán una base sólida. En caso contrario, las conclusiones pueden ser erróneas e, incluso, engañosas.

Tradicionalmente, en los programas de la disciplina Estadística para las diferentes carreras universitarias, incluyendo las humanísticas, se incluyen las principales técnicas “clásicas” tales como las pruebas de hipótesis basadas en la distribución normal de la variable en estudio, análisis de varianza y regresión lineal, estas últimas también incluyendo entre sus supuestos básicos la normalidad de la (s) variable (s) objeto de estudio. De igual modo, los investigadores en dichas ciencias presentan una tendencia general a intentar la aplicación de dichas técnicas al análisis de la información recogida y que sirve de base a la solución de su problema de investigación.

Sin embargo, en dichas ciencias es frecuente, más aún, predominante, la existencia de variables que no corresponden a una escala de intervalo o razón, es decir, cuyos valores no constituyen verdaderamente mediciones, siendo lo más frecuente el trabajo con variables correspondientes al nivel de medida nominal u ordinal. En dicha situación no resultan aplicables las técnicas mencionadas anteriormente, sino otras, incluidas dentro de los llamados “métodos no paramétricos” y otras técnicas tales como la regresión ordinal y logística. El trabajo examina esta cuestión, sugiriendo los temas y técnicas que se consideran más apropiados en carreras e investigaciones de este perfil, describiendo algunas de ellas.

DESARROLLO

"La estadística es una técnica especial apta para el estudio cuantitativo de los fenómenos de masa o colectivo, cuya mediación requiere una masa de observaciones de otros fenómenos más simples llamados individuales o particulares".

Murria R. Spiegel, (1991) dice: "La estadística estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis".

"La estadística es la ciencia que trata de la recolección, clasificación y presentación de los hechos sujetos a una apreciación numérica como base a la explicación, descripción y comparación de los fenómenos". (Yale y Kendal, 1954).

Cualquiera sea el punto de vista, lo fundamental es la importancia científica que tiene la estadística, debido al gran campo de aplicación que posee. La investigación en Psicología, Sociología y Educación, al igual que ocurre en otras ciencias, en buena medida se basa en el manejo de recursos estadísticos como elementos indispensables para llegar a conclusiones aceptables por el resto de la comunidad científica. Dada la peculiaridad de su objeto de estudio, inabordable en la mayoría de los casos si no es a través de perspectivas complejas de relación entre variables, la atención de los investigadores en las ciencias humanas y sociales se concentra cada vez más en la llamada Estadística Multivariante. Los diseños complejos de investigación y análisis, las aportaciones más recientes de la informática para la aplicación de técnicas avanzadas de manipulación de datos y la discusión de estos aspectos desde perspectivas teóricas y aplicadas, preocupan y concentran a multitud de profesionales cuyo quehacer cotidiano es el estudio de cómo se investiga, haciendo de ello su especialidad. Paralelamente, otras especialidades dentro de estas ciencias utilizan el conocimiento ya elaborado y retransmitido, preocupadas más por los resultados y posibilidades que por las condiciones de aplicación y el fundamento de uso, de tal forma que se ha propiciado la utilización de las técnicas estadísticas, sin considerar la adecuación de éstas a las condiciones en las que se aplican.

A su vez, las ciencias sociales se han visto apabulladas en los últimos años por avances vertiginosos en informática y aplicaciones estadísticas, y muy especialmente en la psicología, lo que favorece una absorción de poca calidad por parte de los especialistas en áreas no metodológicas. Por otro lado, la adopción de procedimientos informáticos para realizar tareas metodológicas no parece ser una solución inmediata, considerando la ansiedad que generan los ordenadores, fenómeno muy generalizado.

La fusión de esta creciente complicación de las herramientas de análisis, junto con la discrepancia entre los objetivos de formación y la necesidad de uso de los recursos estadísticos, consigue finalmente que el especialista en áreas aplicadas tienda a descuidar aspectos muy básicos, previos a la aplicación de estos recursos estadísticos complejos. Por otro lado, en muchas ocasiones, la aplicación de herramientas estadísticas se deja arrastrar por hipótesis de comodidad, en el sentido de aplicarse para permitir la ejecución de una prueba o el ajuste de un modelo, no porque son las estrategias más adecuadas, sino porque son las más cómodas.

Entre las técnicas "clásicas" presentes en la casi totalidad de los cursos de Estadística que se imparten para las distintas carreras universitarias se encuentran generalmente los siguientes temas:

I. Estadística Descriptiva

II. Muestreo y Estimación

III. Dótimas de hipótesis

IV. Análisis de Varianza

V. Correlación y Regresión

En el Tema I se acostumbra a tratar las medidas de tendencia central y dispersión más usuales, tales como media aritmética, mediana, moda, varianza, desviación estándar, etc, así como la confección de distribuciones de frecuencias.

Es usual en el Tema II el tratamiento del Muestreo Simple Aleatorio y la estimación puntual y por intervalos de los parámetros de la distribución normal y el parámetro p de la distribución binomial.

Se incluyen habitualmente en el Tema III igualmente pruebas referidas a los parámetros de la distribución normal y la binomial

El tema IV se dedica al problema (que generaliza lo tratado en el anterior) de determinar si la media de una cierta variable es la misma o no en varias poblaciones, determinadas por uno o varios factores cualitativos.

Finalmente el Tema IV trata sobre la relación existente entre variables en una misma población y la elaboración de modelos que relacionen cuantitativamente sus valores.

Como se comentaba en la introducción la insuficiencia fundamental en tales tipos de cursos para carreras humanísticas estriba en el hecho de su limitada aplicación práctica a problemas de la especialidad. En estos son mayoritarias variables cualitativas, a lo sumo correspondientes a una escala ordinal de medición, para las cuales no son aplicables muchas de las herramientas estudiadas.

Escalas de medición

Escala nominal: Sus categorías se representan por valores numéricos u otros símbolos, pero la adopción de un valor u otro por la variable no significa mayor ó menor presencia de una cualidad o característica determinada. Solo es admisible la confección de distribuciones de frecuencias absolutas y la determinación de la moda.

Escala ordinal: Los valores de la variable pueden ser ordenados y cada uno representa una mayor o menor presencia de un atributo o característica dada, pero no existe una unidad de medida, no se conoce la distancia entre dos valores de la escala. Un ejemplo característico lo constituyen las escalas de preferencia u opinión sobre una propuesta dada: “totalmente de acuerdo”, “muy de acuerdo”, “parcialmente de acuerdo”, “en desacuerdo”. Es lícita la confección de distribuciones de frecuencias absolutas y acumuladas, así como el cálculo de la moda y la mediana

Escala de intervalo: Es una escala ordinal con una unidad de medida, se conoce la distancia entre cada valor de la escala, pero no existe un cero absoluto. Pueden servir de ejemplo ilustrativo las escalas de temperatura

Escala de razón: Es una escala de intervalo con un cero absoluto. Variables como longitud, peso, etc serían ejemplos adecuados.

Para las variables en escala de intervalo o razón pueden ser calculadas todas las medidas estadísticas usuales

Así, dentro de la Estadística Descriptiva sería razonable centrar la atención en aquellas medidas apropiadas para escalas a lo sumo ordinal, abundando en medidas como el recorrido y otras generalmente agrupadas en el llamado Análisis Exploratorio de Datos.

De igual forma en los contenidos correspondientes a Dóctimas de Hipótesis, deben tratarse pruebas no paramétricas, como la prueba de la mediana para una población, la dóctima de Fisher, las pruebas de Kolmogorov – Smirnov , Mann – Whitney y Wilconxon, etc.

En lo referente al tema de Análisis de Varianza, esta técnica presupone que la variable aleatoria en estudio; cuya media se desea comprobar si es la misma o no para las distintas poblaciones determinadas por los factores involucrados, posea una distribución normal. Aún cuando la técnica es robusta a la violación moderada de este requisito, no sería correcto aplicarla a variables cuyo nivel de medición corresponde a una escala ordinal.

En este caso correspondería la aplicación de la *prueba de Kruskall – Wallis*.

La prueba permite decidir si **b** muestras independientes proceden o no de la misma población o de poblaciones con medianas iguales.

H_0 : Las **b** muestras proceden de la misma población

H_1 : Las **b** muestras proceden de poblaciones diferentes

Procedimiento:

- Los valores de todas las muestras unidas se ordenan y se les asigna *rangos*
- Se calculan las sumas de rangos de cada muestra

- Se calcula $H = \frac{12}{N(N+1)} \sum_{j=1}^b \frac{R_j^2}{n_j} - 3(N+1)$ donde:

n_j : tamaño de la muestra número **j**

$$N = \sum_{j=1}^b n_j, \text{ o sea, el número total de observaciones}$$

R_j : Suma de los rangos de la muestra **j**

Bajo H_0 el estadígrafo H tiene una distribución aproximadamente $\chi^2(b-1)$ para $b > 3$ y todas las n_j mayores que 5. Por tanto, la región crítica es

R.C : $H > \chi_{\alpha}^2(b-1)$, ya que un mayor valor de H implica mayores diferencias entre las muestras

En efecto, veamos el caso en que las n_j son iguales (todas las muestras del mismo tamaño). Bajo H_0 las sumas de rangos de cada muestra debían ser aproximadamente iguales e iguales a $(1+2+\dots+N)/b = \frac{N(N+1)}{2b}$ de modo que quedaría

$$H = \frac{12}{N(N+1)} \sum_{j=1}^b \frac{R_j^2}{n_j} - 3(N+1) = \frac{12}{N(N+1)} \cdot \frac{b}{N} \left(\frac{N(N+1)}{2b} \right)^2 - 3(N+1) = 0$$

Puesto que los valores de una variable con distribución chi-cuadrado son no negativos este valor 0 es el mínimo posible para H .

Ligas

Al asignar los rangos pueden aparecer valores iguales a los cuáles corresponderían los mismos rangos, en este caso se dice que existen “ligas”. Cuando estas ocurren debe hacerse una corrección, dividiendo el valor de H por $1 - \frac{\sum T}{N^3 - N}$

Donde $T = t^3 - t$, t es el número de observaciones ligadas en un grupo y T se calcula para cada grupo y sus valores se suman

En cuanto al tema de Correlación y Regresión Lineal: existen distintos coeficientes de correlación diseñados para medir el grado de asociación entre variables expresadas en diferentes escalas, tal vez los más pertinentes para este tipo de especialidades serían el **coeficiente de contingencia C** , para dos variables en escala nominal y el **coeficiente de correlación de orden de Spearman**, apropiado para variables en escala ordinal.

En cuanto a la regresión, es decir, la modelación de la relación cuantitativa entre variables para variables nominales y ordinales se presentan como apropiadas dos herramientas específicas básicas: la **regresión logística** y la **regresión ordinal**

Regresión logística

Los modelos de Regresión Logística (RL) se desarrollaron en las últimas décadas del presente siglo para analizar relaciones entre una variable dicotómica y factores cuantitativos y cualitativos potencialmente asociados con ella.

Se trabaja con una variable dependiente con únicamente dos categorías u opciones, en la cual el objetivo es averiguar la probabilidad de ocurrencia de una u otra a partir de una serie de informaciones de una serie de variables que se denominan independientes.

La investigación en Psicología, Sociología y Educación, al igual que ocurre en otras ciencias, en buena medida se basa en el manejo de recursos estadísticos como elementos indispensables para llegar a conclusiones aceptables por el resto de la comunidad científica. Dada la peculiaridad de su objeto de estudio, inabordable en la mayoría de los casos si no es a través de perspectivas complejas de relación entre variables, la atención de los investigadores en las ciencias humanas y sociales se concentra cada vez más en la llamada Estadística Multivariante. Los diseños complejos de investigación y análisis, las aportaciones más recientes de la informática para la aplicación de técnicas avanzadas de manipulación de datos y la discusión de estos aspectos desde perspectivas teóricas y aplicadas, preocupan y concentran a multitud de profesionales cuyo quehacer cotidiano es el estudio de cómo se investiga, haciendo de ello su especialidad. Paralelamente, otras especialidades dentro de estas ciencias utilizan el conocimiento ya elaborado y retransmitido, preocupadas más por los resultados y posibilidades que por las condiciones de aplicación y el fundamento de uso, de tal forma que se ha propiciado la utilización de las técnicas estadísticas, sin considerar la adecuación de éstas a las condiciones en las que se aplican. A su vez, las ciencias sociales se han visto *apabulladas* en los últimos años por avances vertiginosos en informática y aplicaciones estadísticas, y muy especialmente en la psicología, lo que favorece una absorción de poca calidad por parte de los especialistas en áreas no metodológicas. Por otro lado, la adopción de procedimientos informáticos para realizar tareas metodológicas no parece ser una solución inmediata, considerando la ansiedad que generan los ordenadores, fenómeno muy generalizado.

La fusión de esta creciente complicación de las herramientas de análisis, junto con la discrepancia entre los objetivos de formación y la necesidad de uso de los recursos estadísticos, consigue finalmente que el especialista en áreas aplicadas tienda a descuidar aspectos muy básicos, previos a la aplicación de estos recursos estadísticos complejos.

CONCLUSIONES.

Para la enseñanza de métodos estadísticos en las Ciencias Sociales y Humanísticas es necesario tener en cuenta:

- Las técnicas “clásicas” presentes en la casi totalidad de los cursos de Estadística que se imparten para las distintas carreras universitarias.
- Escalas de medición.
- Análisis de Varianza.
- La regresión logística y la regresión ordinal.

BIBLIOGRAFÍA.

1. Freund, J. “Estadística Elemental Moderna”, Editorial Pueblo y Educación, La Habana, 1984. Capítulos I al V.

2. Spiegel, Murray. "Teoría y problemas de Estadística", Editorial Pueblo y Educación, La Habana, 1970.
3. Soler, S. y Suárez, R. "Métodos estadísticos", Editorial Pueblo y Educación, 1987.
4. Statsoft Text Book-Cluster Analysis. Material en la Red Universitaria.
5. Hair, Joseph F. "Análisis Multivariante". Material en la Red Universitaria.
6. Jonson, Dallas E. "Métodos Multivariados Aplicados al Análisis de Datos". Material en la Red Universitaria.