

La estadística, instrumento de investigación científica

Antonio Ramírez Ramírez

Doctor en Educación. Profesor normalista. Ingeniero Industrial. Profesor titular de tiempo completo en la Universidad Pedagógica Nacional, Unidad 141. Profesor de la Normal Superior de Jalisco. Coordinador académico del programa de Doctorado en Administración del Centro de Posgrado en Administración e Informática. Especialista en control de calidad y procesos estadísticos. Asesor de tesis de licenciatura, maestría y doctorado en las áreas de educación, ingeniería y administración. Asesor en el área de calidad del Consejo Técnico Estatal de Educación. Autor de cinco libros en el área de estadística, matemáticas y calidad. Conferencista a nivel nacional y profesor invitado en diversas universidades del país.

Resumen

La estadística, instrumento de investigación científica tiene como propósito fundamental presentar a los investigadores en todas las ramas del quehacer científico, la importancia de la estadística en los procesos de investigación.

Este trabajo muestra cómo la estadística apoya el trabajo del investigador desde el planteamiento del problema, la formulación de la hipótesis y el diseño del tamaño de la muestra, así como las técnicas estadísticas univariantes, bivalentes, multivalentes, y las técnicas estadísticas no paramétricas para el análisis de los datos obtenidos en el trabajo de campo. Asimismo, se hace mención del *software* estadístico disponible para su aplicación.

También se argumenta cómo el método científico encuentra en los métodos estadísticos una herramienta fundamental para alcanzar sus objetivos. Y cómo la estadística aporta información valiosa a los datos obtenidos de un proceso de investigación cualitativa con el fin de obtener conclusiones más asertivas.

Palabras clave: Estadística, investigación científica, ciencia, tecnología.

Desarrollo

La ciencia y la tecnología caracterizan el desarrollo de nuestra sociedad actual, debido a que la ciencia es una actividad social principalmente. El desarrollo de la ciencia y la tecnología ha contribuido a mejorar los niveles de vida, a eliminar epidemias, como: la peste, la polio, la tuberculosis, entre otras enfermedades.

La tecnología es anterior a la ciencia porque ésta requiere instrumentos para su desarrollo, como lo son: los telescopios, el microscopio, los calibradores, los micrómetros, el transportador, el goniómetro, entre otros.

Actualmente la ciencia y la tecnología se presentan en una estrecha relación simbiótica. “La tecnología necesita de la ciencia para seguir desarrollando sus métodos, y la última necesita de la primera para la construcción de sus instrumentos. El vínculo entre una y otra aparece claramente al considerar que los avances científicos y tecnológicos van de la mano: unos son imposibles sin los otros” (Wiesskopf F. V. ,1987:57).

Para Bunge (1987), la ciencia es “conocimiento racional, sistemático, exacto, verificable y, por consiguiente, falible”. La ciencia tiene como objeto buscar continuamente la demostración, el argumento, las razones, las cosas y las pruebas de que lo asentado por ella corresponde con la realidad.

El método científico según, Alvira (1992: 62) “es una referencia y patrón sobre el cual se analizan los problemas y fases de la investigación”.

El método científico está integrado por la:

Teoría.

Hipótesis.

Observación.

Generalizaciones empíricas.

Y se presenta desde dos perspectivas:

- a. Desde el aspecto empírico se refiere al uso de los sentidos, tanto en la observación de los fenómenos como de la experimentación o manipulación física de ellos.
- b. El aspecto racional del método científico se refiere tanto al uso de la razón en el diseño de hipótesis como a la inferencia de las conclusiones. La investigación científica tiene como finalidad la descripción, explicación y predicción de los fenómenos. Una de las herramientas matemáticas fundamentales en la aplicación del método científico es la estadística.

La estadística se define como la aplicación del método científico en el análisis de datos numéricos con el fin de tomar decisiones racionales (Berenson y Levine 1987).

La estadística se divide en:

- a. Estadística descriptiva. Se define como los métodos que implican recopilación, caracterización y presentación de un conjunto de datos con el fin de describir varias de las características.
- b. Estadística inferencial. Se define como aquellos métodos que permiten hacer estimación de una característica de la población o de toma de decisiones con respecto a una población basada solo en los resultados obtenidos de la muestra.

La ciencia y la tecnología requieren de la aplicación del método científico para su desarrollo, y la estadística es una de las herramientas matemáticas fundamentales, apoya al investigador en su proceso de indagación desde la perspectiva de la filosofía analítica en el diseño del:

1. Planteamiento del problema.
2. Operacionalización del problema.
 - 2.1. Formulación de hipótesis.
 - 2.2. Operativización de conceptos teóricos.
 - 2.3. Definición de unidades de análisis.
3. Diseño muestral.
4. Técnicas de análisis de datos.

El planteamiento del problema de investigación. Permite al investigador la descripción de los hechos o situaciones que pueden constituirse en el objeto de estudio para su verificación o comprobación. La identificación del problema de investigación implica formular las variables e indicadores que lo constituyen con el fin de diseñar objetivos e hipótesis de investigación. Un buen planteamiento del problema requiere:

- a. Una relación entre dos o más variables.
- b. El problema y su formulación implica la posibilidad de comprobación empírica.

La estadística aporta al investigador, en su fase de formulación del problema, la definición y clasificación de los tipos de variables aleatorias que conforman la problematización de la investigación. Una variable aleatoria es aquella en la que sus valores numéricos surgen de un proceso de azar o cuando no se puede determinar su valor o no se puede predecir exactamente con anticipación. Un ejemplo de variable aleatoria es el siguiente: a un investigador de mercados le interesa conocer las preferencias de los jóvenes de quince a veinte años, de ambos sexos, respecto a los programas de televisión. Al investigador no le interesa el nombre de los jóvenes, si estudia o no, sino que solamente le

conciernen una variable: la preferencia respecto a los programas de televisión. El análisis estadístico de la información se realiza sobre los atributos de las personas, objetos o entidades. Es necesario, en un problema de investigación, conocer qué tipo de valores tienen las variables en un estudio, porque de esto depende la técnica estadística que se aplique al analizar la información.

Las variables aleatorias se clasifican en dos tipos: variables cualitativas y cuantitativas. Las variables cualitativas reciben el nombre de categóricas. Las variables categóricas se dividen en nominales y ordinales. Las variables nominales permiten al investigador clasificar o ubicar a cada individuo u objeto en una categoría sin importar el orden de los valores de la variable, por ejemplo: el sexo de las personas, su nacionalidad, su origen social, su oficio. Este tipo de variable casi siempre es solicitada por el investigador por medio de un cuestionario.

Las variables categóricas ordinales se caracterizan porque sus valores tienen un orden, por ejemplo: un investigador desea conocer el orden de preferencia de los jóvenes de ambos sexos de los programas de televisión, en este caso los valores que adquiere la variable “programas preferidos” son: las telenovelas, los musicales, la entrevista, las películas y los programas de concursos con las iniciales respectivas: TN, MU, EN, PE, PC. Esta variable “programas de preferencia” clasifica a los jóvenes en cinco categorías.

Otro ejemplo es el siguiente: un investigador social requiere conocer la opinión de los ciudadanos respecto al desempeño de sus gobernantes, les aplica un cuestionario que permita conocer la opinión de las personas y les pregunta: ¿está de acuerdo en que el gobierno ha combatido la delincuencia? Las respuestas son las siguientes:

- Muy en desacuerdo.
- En desacuerdo.
- Ni de acuerdo ni en desacuerdo.
- De acuerdo.
- Muy de acuerdo

Esta variable clasifica a los ciudadanos que contestaron la pregunta según la opción elegida.

Las variables cuantitativas se denominan numéricas y se clasifican en discretas y continuas. Las variables numéricas discretas son aquellas que solo pueden asumir ciertos valores dentro de un intervalo, surgen de un proceso de conteo y solamente pueden tomar valores en la escala de números enteros. Ejemplos de variables numéricas discretas son: el número de mesas y sillas que se encuentran en un salón de clases, el número de accidentes automovilísticos en una ciudad determinada por día, el número de estudiantes aceptados en la universidad en un semestre.

Las variables numéricas continuas asumen teóricamente cualquier valor en algún intervalo de números reales. Ejemplo de este tipo de variables son: la estatura y el peso de las personas.

Otro proceso importante en la operacionalización de las variables es el proceso de medición. Se entiende por medición “la asignación de números a objetos según determinadas reglas (Glass y Stanley: 1980). “Es el proceso de vincular conceptos abstractos a indicadores empíricos” (Carmines y Zeller, 1979). A continuación se presenta un esquema de las escalas de medición utilizadas en un proceso de investigación.

Características de medición			
Escala nominal	Se clasifican los objetos y las clases se enumeran.	Ejemplos: Estado civil. Sexo. Lugar de nacimiento	Técnicas estadísticas. Porcentajes. Moda. Ji cuadrada.
Escala ordinal	Establece una relación ordenada entre objetos y acontecimientos.	Escala brinell Calificaciones escolares. Rangos militares. Orden en las preferencias.	Percentiles. Mediana. Correlación por rango-orden.
Escala de intervalo	Permite colocar los datos en una escala numérica continua, en la cual los intervalos entre las partes son iguales. El punto cero de la escala de intervalo es arbitrario y no refleja la ausencia del atributo.	La medición de la temperatura. Actitudes. Opiniones.	Rango. Media. Desviación estándar. Correlación de Pearson...
Escala de razón	Tienen todas las propiedades de una escala de intervalo. Una puntuación cero indica la ausencia de la propiedad que se mide.	Altura. Peso. Temperatura en una escala de grados <i>kelvin</i> . Edades.	Todas las técnicas estadísticas paramétricas.

La operacionalización del problema se realiza desde dos vertientes: la teórica-analítica (concepto, constructos y variables de estudio) y la vertiente poblacional que implica el tipo de población en estudio y el diseño del tamaño de la muestra representativa. Este proceso de la operacionalización del problema de investigación comprende la formulación de hipótesis, la operacionalización de conceptos teóricos y la definición de unidades de análisis.

Las hipótesis son explicaciones previas acerca de las relaciones entre dos o más variables sustentadas en la teoría y sujetos a su comprobación empírica. Una hipótesis estadística es una proposición sobre los parámetros de una o más poblaciones.

Una prueba de hipótesis es un procedimiento que conduce a una decisión sobre una hipótesis determinada. El proceso de la prueba de hipótesis depende del uso de la información obtenida en una muestra aleatoria de la población objeto de estudio. Si esta información es consistente con la hipótesis se concluye que esta es verdadera, en cambio, si la información es inconsistente con la hipótesis, se concluye que es falsa. Es necesario destacar que la verdad o falsedad de una hipótesis nunca puede conocerse con certidumbre, al menos que pueda analizarse a toda la población objeto de estudio.

En un proceso de prueba de hipótesis estadísticas se establece la hipótesis nula y la hipótesis alternativa. La hipótesis nula siempre expresa el estado actual en la que asegura que no hay diferencias. Esta hipótesis se identifica con el símbolo H_0 . La hipótesis alternativa se establece como el opuesto de la hipótesis nula y representa la conclusión que se apoya si la hipótesis nula se rechaza, y se identifica con el símbolo H_1 .

La estadística inferencial cuenta con una serie de pruebas de hipótesis, las más utilizadas en la investigación científica son:

1. Prueba de hipótesis con una muestra.
 - 1.1. Prueba de hipótesis Z para la media.
 - 1.2. Prueba de hipótesis "t" para la media.
 - 1.3. Prueba de hipótesis para la proporción.
2. Prueba con dos muestras y "c" muestras de datos numéricos.
 - 2.1. Comparación de dos muestras independientes: prueba "t" para las diferencias entre dos medias.
 - 2.2. Prueba F para las diferencias en dos varianzas.
 - 2.3. Comparación de dos muestras relacionadas: prueba "t" para la diferencia en la media.
 - 2.4. Modelo totalmente aleatorio; análisis de varianza de un factor.

La operacionalización de conceptos teóricos consiste en pasar las variables generales conceptuales a indicadores o índices. Su objetivo es transformar el concepto teórico en un observable. Estos indicadores o índices se expresan con variables categóricas o numéricas, las cuales determinan la escala de medición correspondiente.

Con relación a la definición de unidades de análisis o de observación, se define primero el concepto de población como "un conjunto de unidades en las que se desea obtener cierta información" (Sánchez Crespo, 1971). Las unidades de análisis u observación pueden ser: personas, familias, escuelas, tornillos y demás objetos o personas.

Para obtener el tamaño de la muestra de una población determinada se requiere tomar en cuenta los siguientes factores:

1. El tiempo y los recursos disponibles.
2. Las técnicas de muestreo.
3. Las técnicas de análisis de datos.
4. La estimación de la varianza.
5. El margen de error máximo posible.
6. El nivel de confianza.

Existen modelos matemáticos para obtener el tamaño de la muestra, en este caso solo se presentan las más utilizada.

$$n = \frac{Z^2 \sigma^2}{e^2}$$

n = tamaño de la muestra

Z = intervalo de confianza.

e = error de muestro aceptable.

σ = desviación estándar.

Esta formula se utiliza cuando se desconoce el tamaño de la población.

Tamaño de la muestra para estimar proporciones poblacionales. Cuando el muestreo se hace en una población infinita o en una población finita con reemplazamiento.

$$n = \frac{Z^2 P(1-P)}{e^2}$$

n = tamaño de la muestra

Z = nivel de confianza.

e = error de muestreo aceptable.

P = proporción de la población que tiene la característica de interés.

Cuando el muestreo se realiza sin reemplazamiento en una población finita se utiliza la siguiente formula:

$$n = \frac{NZ^2 Pq}{(N-1)e^2 + Z^2 Pq}$$

n = tamaño de la muestra.

N = tamaño de la población.

p = la proporción de la población que tiene la característica de interés.

q = la proporción de la población que no tiene la característica de interés.

e = error de muestreo aceptable.

Z = nivel de confianza deseado.

Al terminar el trabajo de campo y la recogida de la información se inicia la fase de análisis e interpretación de los datos. Actualmente para esta etapa del proceso de investigación se cuenta con paquetes estadísticos, los más comunes son: *Statistical Package For The Social Sciences (SPSS)*, *el Statistical Analysis System (SAS)*, *el Biomedical Computer Programs (BMPD)*, *el Minitab* y *el Statgraphics*.

Entre las técnicas que se usan para sistematizar los datos obtenidos del trabajo de campos se tiene: el diagrama de barras, el histograma, las gráficas *Steam and Leaf*, el diagrama de caja, el polígono de frecuencia, la ojiva, la gráfica de sectores, entre otros. Con el fin de describir cómo se agrupan los atributos de una variable alrededor de un valor típico, se tienen las medidas denominadas de tendencia central, las más utilizadas son: la media aritmética, la mediana y la moda.

Una vez que se ha calculado el valor típico de una serie de datos, el investigador desea saber el grado en que los valores difieren de este valor típico. Para esta situación se utiliza el término de dispersión para describir el grado en que una serie de valores varían respecto a este valor típico. Las medidas de dispersión más comunes son: rango, varianza, desviación estándar y el coeficiente de variación. El investigador también suele utilizar medidas de tendencia no central como los cuartiles, los cuales representan valores que dividen a la distribución en partes iguales, las más utilizadas son:

Los cuartiles dividen la distribución en cuatro partes iguales.

Los deciles en diez partes iguales.

Los percentiles en cien partes.

El análisis bivariante implica la relación entre dos variables. Las técnicas más comunes son el:

1. Análisis de regresión lineal simple.
2. Medidas de variación en la regresión.
3. El coeficiente de determinación.
4. Error estándar de la estimación.
5. Análisis residual.
6. Prueba "t" para la pendiente.
7. Prueba F para la pendiente.
8. Estimación de intervalos de confianza.
9. Coeficiente de correlación.

El análisis multivariable permite el análisis de más de dos variables en una muestra de observaciones. Las técnicas del análisis multivariante se clasifican en técnicas de dependencia y técnicas de interdependencia. Las técnicas de dependencia permiten al investigador valorar el grado de relación entre variables dependientes e independientes. Las técnicas más utilizadas son:

1. Regresión múltiple.
2. Análisis conjunto.
3. Análisis discriminante.
4. Análisis multivariante de la varianza.
5. La regresión logística.
6. Correlación canónica.

Las técnicas de interdependencia permiten al investigador identificar las estructuras entre una serie de variables, observaciones u objetos definidos. La identificación de las estructuras entre las observaciones o varias variables permite establecer nuevas relaciones entre ellas, así como analizar la existencia, asociación o relación mutua entre variables, sin establecer la diferencia entre las dependientes e independientes, las más utilizadas son:

1. Análisis factorial.
2. Análisis cluster.
3. Análisis multidimensional.

También el investigador tiene a su disposición las técnicas de estadística no paramétricas para el análisis de los datos. Las pruebas no paramétricas toman solamente en cuenta si ciertos puntajes son más altos o más bajos que otros. Una ventaja de estas técnicas es que pueden usarse cuando los datos experimentales se pueden medir solo a nivel ordinal, es decir, cuando solamente se pueden establecer rangos en orden de magnitud. Además se aplican a muestras pequeñas (se considera muestras pequeñas a las menores de 30 datos). Las pruebas no paramétricas más comunes son:

1. Prueba de *Wilcoxon* para dos condiciones (datos ordinales y diseños relacionados).
2. Prueba de *Mann-Whitney* para diseños no relacionados con dos condiciones.
3. La prueba de Friedman, cuando se utiliza tres o más condiciones en diseños relacionados.
4. La prueba de tendencia *L de Page*. Cuando se examina la tendencia en tres o más condiciones en un diseño relacionado.
5. La prueba de *Kruskal-Wallis*. Cuando se tienen tres o más condiciones en diseños no relacionados.
6. La prueba de *Tendencia de Jonckheere*, se utiliza cuando se examina las tendencias en tres o más condiciones. Se utiliza para diseños no relacionados.

Todas las técnicas estadísticas mencionadas se pueden aplicar utilizando el *software* estadístico en el análisis de los datos obtenidos del trabajo de campo.

Es indudable que los métodos estadísticos paramétricos y no paramétricos aportan a la labor del investigador una herramienta valiosa en la aplicación del método científico. El método se caracteriza por:

1. La verificación del conocimiento puede lograr el establecimiento de la asociación entre variables por medio de las pruebas de hipótesis.
2. El relativismo del conocimiento. Los resultados de una prueba estadística solo son válidos para datos contenidos en la muestra y su posible generalización probabilística a la población de donde se obtuvo la muestra.

3. Acumulación de conocimiento. El conocimiento científico se desarrolla tomando en cuenta el conocimiento previamente construido, por lo que cada nueva investigación utiliza la estructura técnica y los datos estadísticos ya existentes.
4. Explicación y predicción en el conocimiento. El conocimiento científico aporta la posibilidad de explicar los hechos, fenómenos y/o realidad. Tales explicaciones forman parte de la teoría. La estadística por medio de sus técnicas inferenciales y el diseño experimental coadyuva a la explicación y predicción de los fenómenos naturales y sociales.
5. Aplicación del conocimiento. La utilidad práctica de la ciencia es posible por medio de la tecnología, la cual construye instrumentos para la solución de problemas de la vida social y natural. En este aspecto la estadística también contribuye con la aplicación de los métodos estadísticos en todas las áreas de la ciencia y de la actividad humana en general, por ejemplo: la mercadotecnia, la contabilidad, la ingeniería, la medicina, la educación, la sociología y la mejora constante de los procesos industriales para lograr la calidad total, la confiabilidad, la productividad y rentabilidad. Además en la aplicación de los procesos administrativos como lo son: la reingeniería, *benchmarking*, entre otros.

Conclusión

Es necesario que el investigador de las ciencias sociales y naturales aplique las técnicas estadísticas en sus procesos de investigación con el fin de obtener mejores conclusiones al complementarlos con los datos obtenidos de la aplicación de los métodos de la investigación cualitativa.

Bibliografía

Berenson, M et al

2001 Estadística *para administración*. México. Pearson Educación.

Camacho, J.

2003 Estadística *con SPSS para Windows*. México. Alfaomega.

Catena, A. et al.

2005 *Análisis multivariado*. Madrid. Biblioteca Nueva.

Cea, M. A.

2001 *Metodología cuantitativa*. Madrid. Síntesis. Sociología.

Greene, J. y M. D Oliveira.

1984 *Pruebas estadísticas para psicología y ciencias sociales*. Bogotá. Norma.

Hernández, R. et al.

2006 Metodología *de la investigación*. México. Mc Graw Hill.

Mardones J. 1988

Filosofía de las ciencias humanas y sociales. México. Fontamara.

Maxim, P. 2002 *Métodos cuantitativos aplicados a las ciencias sociales*. México. Oxford.