
La Estadística Social como Herramienta Fundamental en la Investigación Científica: Fundamentos y Aplicaciones Descriptivas

Este documento explora la relevancia de la estadística social como una disciplina esencial para la investigación en ciencias sociales, detallando su propósito, su subordinación al método científico y sus principales ramas. Se profundiza en la estadística descriptiva, explicando sus medidas de tendencia central (moda, mediana y media aritmética) y las diversas formas de organizar y presentar datos, como tablas de frecuencias y gráficos. El objetivo es proporcionar una comprensión sólida de cómo la estadística permite transformar datos brutos en información significativa, facilitando la interpretación de fenómenos sociales y la toma de decisiones informadas.

Alexis E. Rebolledo Carreño

Metodólogo

[ORCID: 0000-0003-2728-5723](https://orcid.org/0000-0003-2728-5723)

19 de mayo de 2026

Introducción a la Estadística Social en la Investigación Científica

La estadística constituye una disciplina fundamental que provee las herramientas metodológicas necesarias para la recolección, organización, resumen, análisis e interpretación de datos, permitiendo la realización de inferencias válidas a partir de ellos (Hernández-Sampieri et al., 2014). En el ámbito de la investigación social, su aplicación es crucial, ya que facilita la comprensión de fenómenos complejos y la validación de hipótesis sobre el comportamiento humano y las estructuras sociales. El propósito primordial de la estadística en este contexto es transformar conjuntos de datos brutos en información significativa y accesible, lo cual es indispensable para la construcción de conocimiento empírico y la formulación de políticas públicas basadas en evidencia.

El uso de la estadística en el marco de una investigación se subordina intrínsecamente a los objetivos planteados y al nivel de medición de las variables involucradas, pero, de manera aún más crítica, a las hipótesis de trabajo que guían el estudio (Creswell & Creswell, 2018). Esto significa que la elección de las técnicas estadísticas adecuadas no es arbitraria, sino que debe estar en consonancia con la naturaleza de las preguntas de investigación y las suposiciones teóricas. La estadística es, por excelencia, una estrategia metodológica inherente a las investigaciones de corte cuantitativo, donde la medición y el análisis numérico son centrales, y también desempeña un papel vital en aquellos estudios que adoptan una integración metodológica, como los diseños mixtos (Tashakkori & Teddlie, 2010).

Dentro de la estadística, se distinguen tradicionalmente dos ramas principales que cumplen funciones complementarias en el proceso investigativo: la estadística descriptiva y la estadística inferencial. La primera se encarga de organizar y resumir los datos, mientras que la segunda permite generalizar los hallazgos de una muestra a una población más amplia, estimar parámetros y probar hipótesis (Field, 2018). Ambas ramas son esenciales para proporcionar una comprensión completa de los datos y para extraer conclusiones robustas que contribuyan al avance del conocimiento en las ciencias sociales.

La Subordinación de la Estadística al Método Científico

La aplicación de la estadística en cualquier campo de estudio, y especialmente en las ciencias sociales, debe estar rigurosamente subordinada a los principios y etapas del método científico. Esto implica que la estadística no es un fin en sí misma, sino una herramienta poderosa que sirve para apoyar cada fase del proceso de investigación (Bunge, 2012). Desde la concepción inicial de un problema hasta la comunicación de los resultados, las técnicas estadísticas proporcionan el andamiaje necesario para garantizar la objetividad y la validez de los hallazgos.

El ciclo del método científico, en el que la estadística se inserta, comienza con el planteamiento de hipótesis claras y verificables que guían la investigación. Posteriormente, se procede al diseño de la investigación, donde se define la estrategia metodológica, la población de estudio, el muestreo y los instrumentos de recolección de datos. Una vez recolectados, los datos son analizados mediante técnicas estadísticas apropiadas, lo que permite la verificación o refutación de las hipótesis iniciales. Finalmente, a partir de este análisis, se obtienen conclusiones que contribuyen a la teoría existente y pueden generar nuevas preguntas de investigación, cerrando así el ciclo virtuoso del conocimiento científico (Popper, 1959).

Estadística Descriptiva: Organización y Síntesis de Datos

La estadística descriptiva es la rama de la estadística que se ocupa de la recolección, organización, resumen y presentación de un conjunto de datos (Babbie, 2021). Su objetivo principal es 'decir algo' de la población o muestra que se está estudiando, permitiendo al investigador acceder con mayor facilidad a las variabilidades y características inherentes a los datos. A través de diversas técnicas, la estadística descriptiva centraliza y organiza los datos de manera que sean comprensibles, facilitando la identificación de patrones, tendencias y anomalías sin intentar hacer inferencias más allá de los datos observados.

Cuando los datos se encuentran en su 'estado natural', es decir, sin procesar, resulta difícil

visualizar su regularidad, predominio, variabilidad o recurrencia. Por ello, se requiere la implementación de estrategias de organización numérica que permitan la centralización y la síntesis de esta información. La estadística descriptiva ofrece un conjunto de medidas y representaciones gráficas diseñadas para este propósito, transformando grandes volúmenes de datos en resúmenes concisos y comprensibles que son la base para cualquier análisis posterior, ya sea descriptivo o inferencial (Field, 2018).

Medidas de Tendencia Central

Las medidas de tendencia central son valores numéricos que describen el centro de un conjunto de datos, proporcionando una idea de dónde se agrupan la mayoría de las observaciones. Estas medidas son fundamentales para resumir la información y ofrecer una representación concisa de las características principales de una distribución (Kerlinger & Lee, 2002). La estadística descriptiva nos entrega tres medidas de centralización de datos ampliamente utilizadas: la moda, la mediana y la media aritmética, cada una con sus propias propiedades y aplicaciones específicas dependiendo del tipo de variable y la forma de la distribución de los datos.

La Moda (Mo)

La moda es la medida de tendencia central más sencilla de identificar, representando el valor o categoría que aparece con mayor frecuencia en un conjunto de datos (Hernández-Sampieri et al., 2014). En la observación sociocientífica, es posible detectar datos recurrentes que reflejan comportamientos, opiniones, actividades o preferencias culturales. La representación numérica de estos patrones se expresa a través de la identificación de la moda, que es el dato que más se repite en la muestra. Su principal ventaja radica en su aplicabilidad a variables de cualquier nivel de medición, incluyendo las nominales, ordinales y de intervalo/razón.

La moda es particularmente útil para variables categóricas, donde no se pueden calcular promedios numéricos. Por ejemplo, al analizar el color de ojos más común en una población o el estilo musical preferido por un grupo de jóvenes, la moda proporciona una indicación clara

de la categoría dominante. Aunque es una medida robusta ante valores extremos, su principal limitación es que puede haber varias modas (distribuciones bimodales o multimodales) o ninguna, y no utiliza toda la información disponible en los datos numéricos (Field, 2018). Sin embargo, su simplicidad y versatilidad la hacen indispensable en el análisis descriptivo inicial.

La Mediana (Mdn)

La mediana es el valor central en un conjunto de datos ordenados, que divide la distribución en dos partes iguales, de modo que el 50% de las observaciones son menores o iguales a ella y el 50% son mayores o iguales (Babbie, 2021). En una distribución de elementos, datos o frecuencias, la mediana es un punto de referencia intermedio que permite separar la muestra en dos mitades. Para calcularla, los datos deben estar ordenados de forma ascendente o descendente. Si el número de observaciones (N) es impar, la mediana es el valor que se encuentra en la posición $(N+1)/2$. Si N es par, la mediana es el promedio de los dos valores centrales.

La mediana es una medida de tendencia central robusta frente a la presencia de valores atípicos o extremos, lo que la hace particularmente útil en distribuciones asimétricas donde la media podría ser engañosa (Field, 2018). Se utiliza comúnmente con variables de nivel ordinal y de intervalo/razón, ya que requiere que los datos puedan ser ordenados. Por ejemplo, al analizar ingresos o tiempos de respuesta, la mediana puede ofrecer una representación más precisa del 'valor típico' que la media, si existen unos pocos valores extremadamente altos o bajos que distorsionen el promedio.

La Media Aritmética ($\bar{x}$)

La media aritmética, comúnmente conocida como promedio, es la medida de tendencia central más utilizada y familiar. Se calcula sumando todos los valores de un conjunto de datos y dividiendo el resultado por el número total de observaciones (Hernández-Sampieri et al., 2014). Una buena manera de reconocer una característica general de una muestra es a través de la identificación de promedios, ya que la media representa el punto de equilibrio de la

distribución. Su fórmula general para datos desagrupados es $\bar{x} = \Sigma x / N$, donde Σx es la suma de todos los valores y N es el número total de observaciones.

Para datos agrupados en tablas de frecuencias, la media aritmética se calcula como $\bar{x} = \Sigma (f * x) / N$, donde 'f' es la frecuencia de cada valor o punto medio de intervalo de clase, y 'x' es el valor o punto medio. La media es sensible a todos los valores de la distribución y es el centro de gravedad de los datos, lo que la hace ideal para análisis estadísticos más avanzados. Sin embargo, su principal desventaja es que es muy susceptible a la influencia de valores extremos o atípicos, que pueden distorsionar su representatividad en distribuciones asimétricas (Field, 2018). Por esta razón, es fundamental considerar la forma de la distribución antes de decidir si la media es la medida de tendencia central más apropiada.

Ejemplos de Cálculo de Medidas de Tendencia Central

Para ilustrar la aplicación de estas medidas, consideremos un ejemplo hipotético de minutos de atraso de estudiantes. Supongamos que se registran los minutos de atraso de 32 estudiantes en un curso. Si los datos brutos son 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 6, 7, 8, 9, 9, 9, 10, 10. Ordenando estos datos, podemos calcular la moda, mediana y media. La moda sería 1, ya que es el valor que más se repite (11 veces). La mediana, al ser $N=32$ (par), se encontraría entre la posición 16 y 17. Si los valores en esas posiciones son 3 y 3, la mediana sería 3. La media aritmética se calcularía sumando todos los minutos de atraso y dividiendo por 32.

Un ejemplo más detallado de cálculo para datos desagrupados podría ser: Si los minutos de atraso de 32 estudiantes son: 1 (11 veces), 2 (5 veces), 4 (3 veces), 5 (2 veces), 6 (1 vez), 8 (1 vez), 9 (3 veces), 10 (6 veces). La suma total de los minutos de atraso sería $(1 \times 11) + (2 \times 5) + (4 \times 3) + (5 \times 2) + (6 \times 1) + (8 \times 1) + (9 \times 3) + (10 \times 6) = 11 + 10 + 12 + 10 + 6 + 8 + 27 + 60 = 144$. Por lo tanto, la media aritmética sería $144 / 32 = 4.5$ minutos. En este caso, la moda es 1, la mediana es 3 y la media es 4.5, lo que sugiere una distribución asimétrica con una concentración de valores bajos y algunos valores más altos que elevan el promedio.

Para datos agrupados, consideremos una tabla de frecuencias donde 'x' representa el valor

de la variable y 'f' su frecuencia. Por ejemplo, si tenemos los siguientes datos: $x=1$ ($f=0$), $x=2$ ($f=3$), $x=3$ ($f=5$), $x=4$ ($f=4$), $x=7$ ($f=2$), $x=8$ ($f=1$), con un N total de 15. La moda sería 3, ya que tiene la mayor frecuencia ($f=5$). Para la mediana, $(N+1)/2 = (15+1)/2 = 8$, por lo que la mediana sería el valor en la octava posición de los datos ordenados. Si los datos ordenados son 2,2,2,3,3,3,3,3,4,4,4,4,7,7,8, la mediana es 3. La media se calcularía como $\Sigma(f \cdot x) / N = (0+6+15+16+14+8) / 15 = 59 / 15 = 3.93$. Estos ejemplos demuestran cómo las medidas de tendencia central proporcionan una visión rápida y concisa de la distribución de los datos.

Presentación Ordenada de Datos: Tablas y Gráficos

Una vez recolectados, los datos brutos carecen de significado inmediato y requieren ser organizados para su análisis. Las tablas de frecuencias y las representaciones gráficas son dos maneras equivalentes y complementarias de presentar la información de forma ordenada y comprensible (Miles, Huberman & Saldaña, 2014). Ambas herramientas permiten al investigador y a la audiencia visualizar la distribución de las variables, identificar patrones y resumir las características clave de la muestra o población estudiada, sin perder la esencia de la información recogida.

Tablas de Frecuencias

Las tablas de frecuencias son una herramienta fundamental en la estadística descriptiva para organizar y resumir datos, exponiendo la información recogida en la muestra de tal manera que se minimiza la pérdida de información (Babbie, 2021). Estas tablas permiten visualizar la distribución de una variable a través de diferentes tipos de frecuencias. Las frecuencias absolutas (f) contabilizan el número de individuos o casos que pertenecen a cada modalidad o categoría de la variable. Las frecuencias relativas (fr), expresadas a menudo como porcentajes, indican la proporción de individuos en cada categoría respecto al total, facilitando la comparación entre diferentes conjuntos de datos.

Además, las tablas de frecuencias pueden incluir frecuencias acumuladas, las cuales solo tienen sentido para variables de nivel ordinal y numéricas (de intervalo o razón). Las

frecuencias acumuladas muestran el número o porcentaje de observaciones que caen en una categoría dada o en cualquier categoría anterior. Esto es particularmente útil para responder preguntas como '¿qué porcentaje de individuos tiene menos de X hijos?' o '¿cuántos individuos tienen una puntuación igual o inferior a Y?'. Por ejemplo, si el 83.8% de los individuos tiene menos de 3 hijos, esta información se obtiene directamente de la frecuencia acumulada, ofreciendo una visión rápida de la distribución de la variable (Field, 2018).

Consideremos un ejemplo simple de datos desordenados y ordenados en tablas. Si tenemos una muestra de género (M, H, H, M, M, H, M, M, M, H), podemos organizar estos datos en una tabla de frecuencias. La variable es 'Género' con modalidades 'Hombre' (H) y 'Mujer' (M). Contando, obtenemos 4 hombres y 6 mujeres, para un tamaño muestral total de 10. La tabla mostraría: Hombre con una frecuencia de 4 (40%) y Mujer con una frecuencia de 6 (60%). Esta organización simplifica la interpretación de la distribución del género en la muestra, transformando una lista de caracteres en información estructurada y comprensible.

Gráficos para Variables Categóricas

Las representaciones gráficas son una forma visual y efectiva de presentar la información contenida en las tablas de frecuencias, facilitando la comprensión rápida de las distribuciones de datos (Miles, Huberman & Saldaña, 2014). Para variables categóricas, como el género o el nivel de felicidad, los diagramas de barras y los diagramas de sectores son las opciones más comunes y adecuadas. Estos gráficos permiten visualizar la magnitud relativa de cada categoría y comparar las frecuencias o porcentajes de manera intuitiva.

Los diagramas de barras representan cada categoría de la variable con una barra, cuya altura es proporcional a la frecuencia (absoluta o relativa) de esa categoría. Son particularmente útiles para comparar las frecuencias de diferentes categorías y pueden aplicarse tanto a variables nominales como ordinales, e incluso a variables discretas. Por otro lado, los diagramas de sectores, también conocidos como gráficos de pastel o torta, dividen un círculo en 'sectores' o 'rebanadas', donde el área de cada sector es proporcional a la frecuencia relativa de la categoría que representa. Estos gráficos son excelentes para mostrar la

proporción de cada categoría respecto al total, siendo más efectivos cuando el número de categorías es limitado (Field, 2018). Ambas herramientas visuales complementan las tablas de frecuencias al ofrecer una perspectiva gráfica que a menudo es más accesible que los datos numéricos puros.

Conclusión

La estadística social se erige como una disciplina indispensable en el panorama de la investigación científica, proporcionando un marco riguroso para la comprensión y el análisis de los fenómenos sociales. Su propósito trasciende la mera manipulación de números, al ofrecer una estrategia sistemática para la recolección, organización, resumen y análisis de datos, culminando en la capacidad de realizar inferencias significativas. La subordinación de la estadística al método científico asegura que cada paso, desde la formulación de hipótesis hasta la obtención de conclusiones, se realice con la máxima objetividad y validez, consolidando así el conocimiento empírico en las ciencias sociales (Hernández-Sampieri et al., 2014; Bunge, 2012).

En particular, la estadística descriptiva desempeña un papel crucial al transformar conjuntos de datos brutos en información estructurada y comprensible. Las medidas de tendencia central —moda, mediana y media aritmética— ofrecen perspectivas complementarias sobre el 'centro' de una distribución, mientras que las tablas de frecuencias y los gráficos visualizan la distribución de las variables, revelando patrones y características esenciales. Estas herramientas no solo facilitan la interpretación inicial de los datos, sino que también sientan las bases para análisis inferenciales más complejos, permitiendo a los investigadores sociales abordar preguntas de investigación con mayor profundidad y rigor metodológico (Field, 2018; Babbie, 2021). En suma, la estadística es una aliada insustituible para la generación de conocimiento robusto y la toma de decisiones informadas en el ámbito social.

Referencias bibliográficas

- Babbie, E. (2021). *The practice of social research* (15th ed.). Cengage Learning.
- Bunge, M. (2012). *La ciencia, su método y su filosofía*. Editorial Sudamericana.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill Education.
- Kerlinger, F. N., & Lee, H. B. (2002). *Foundations of behavioral research* (4th ed.). Wadsworth/Thomson Learning.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage Publications.
- Popper, K. R. (1959). *The logic of scientific discovery*. Routledge.
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Handbook of mixed methods in social and behavioral research* (2nd ed.). Sage Publications.

Cómo citar este documento

Alexis E. Rebolledo Carreño (2026). *La Estadística Social como Herramienta Fundamental en la Investigación Científica: Fundamentos y Aplicaciones Descriptivas*. Documento de trabajo metodológico. ORCID: <https://orcid.org/0000-0003-2728-5723>