

## *Aplicaciones estadísticas en las Ciencias Sociales*



*Juan Camacho Rosales*

Profesor Titular de Metodología de las Ciencias del Comportamiento  
Departamento de Didáctica e Investigación Educativa y del Comportamiento  
Universidad de La Laguna

### *La Estadística Matemática*

A medida que la Ciencia progresa, sus teorías se van haciendo más y más matemáticas en la forma. Hay una relación positiva entre el progreso de una Ciencia y el grado de su desarrollo matemático.

No es necesario que el investigador en Ciencias Sociales sea un especialista en áreas matemáticas concretas, lo verdaderamente importante es que sepa acercarse con mentalidad matemática a los problemas que se le plantean. La mentalidad matemática se define como la comprensión del proceso lógico subyacente al razonamiento matemático: entender la estructura formal del modelo matemático y las condiciones que lo hacen posible.

Tiene que haber un compromiso de manera que se simplifique la realidad concreta lo menos posible, pero, a la vez, lo suficiente para que el modelo creado a partir de la realidad sea fácilmente manejable desde el punto de vista instrumental matemático.

Es necesario una buena cierta matemática para conocer la potencia y debilidad de las técnicas estadísticas y consiguientemente para saber usarlas con eficacia y la vez con prudencia.

Para estudiar Estadística Matemática se necesita cálculo avanzado y álgebra de matrices, sin embargo tal madurez no es indispensable para comprender las bases de la Estadística Aplicada.

El sentimiento de satisfacción y tranquilidad que resulta de dominar un lenguaje lógico y no ambiguo compensa la ocasional ansiedad que se desencadena al descubrir que se ha expresado un absurdo explícitamente y a todas luces.

Desde un punto de vista matemático muchas de nuestras afirmaciones están incompletas, mal encuadradas o son imprecisas. Pero, por otro lado, muchas de estas ideas pueden ser entendidas intuitivamente, y es mejor una comprensión intuitiva que ninguna comprensión en absoluto.

Yo prefiero que los ejemplos que se usen en la enseñanza sean hipotéticos, porque es más importante tener un problema simple y plausible que el estudiante pueda comprender y que ilustre el método claramente que otro que simplemente asombre al estudiante con nuestra sabiduría.

Las principales aplicaciones estadísticas en cualquier campo, no sólo el de las Ciencias Sociales, descansan sobre el hecho de poder hacer observaciones o experimentos repetidos, esencialmente, bajo las mismas condiciones. En algunas áreas de la investigación, los objetos o fenómenos observados bajo las mismas condiciones variarán sólo en pequeña medida (en las ciencias físicas, donde las observaciones controladas dan prácticamente los mismos resultados). Pero, por otro lado, especialmente en las Ciencias Sociales, aunque el experimentador haga un esfuerzo sobrehumano para observar repetidamente bajo las mismas condiciones, se encontrarán diferencias entre las observaciones y las diferencias, ordinariamente, no serán despreciables.

La Estadística Matemática es una teoría acerca de la incertidumbre, la tendencia de los resultados a variar cuando observaciones repetidas se hacen bajo condiciones idénticas.

La Estadística es el estudio de fenómenos donde, bajo un mismo conjunto de condiciones, las medidas obtenidas presentan variabilidad, y por tanto resultados impredecibles a priori; es decir, existe incertidumbre asociada al conocimiento del objeto de estudio. Aceptado que la Estadística trata sobre la incertidumbre, cabe preguntarse si la naturaleza está determinada o, en realidad, la incertidumbre es inherente a la misma, y por tanto está indeterminada. Y si está indeterminada entonces la Estadística tratará sobre la misma esencia de la realidad empírica.

Definamos entonces la Estadística como aquella manera de pensar de la cual se deriva una forma de representar los sistemas y razonar sobre ellos, sobre una naturaleza que se muestra indeterminada. La Estadística puede considerarse una Ciencia que guía la extracción de conocimiento, e implica una manera de conceptuar cualquier problema donde la incertidumbre es inherente a la comprensión del objeto de estudio y, por lo tanto, nuestro discernimiento sólo puede ser probabilístico y expresado mediante leyes estadísticas.

Aunque la organización de la información, las transformaciones y la depuración de los datos no sean características esenciales de la Estadística, eso no implica que no puedan ser incluidas en una definición de la disciplina.

El objetivo de la Estadística como Ciencia es mejorar el nivel de vida de la sociedad. Estadística deriva de la palabra Estado, y etimológicamente significa recoger información para tomar decisiones de cómo repartir comida o trabajo.

La Estadística moderna se ocupa de la recolección, análisis e interpretación de información, tanto cuantitativa como cualitativa. Y los métodos estadísticos son particularmente útiles cuando hay variabilidad en la medición.

### ***Utilidad de la Estadística en las Ciencias Sociales***

Un estadístico trabajando en el campo de las Ciencias Sociales se ocupa de las siguientes cuestiones:

- ¿qué datos se necesita recoger?
- ¿cómo se pueden usar los recursos disponibles más eficientemente para recolectar los datos?
- ¿cómo especificar un modelo matemático que describa el proceso que ha generado los datos?
- depuración y transformación de los datos
- ¿cómo presentar los datos de manera que transmitan sus rasgos más esenciales de una manera clara?
- ¿qué conclusiones se pueden extraer de los datos y cuál es el grado de incertidumbre de estas conclusiones?
- ¿qué acciones se deben tomar en base a las conclusiones extraídas de los datos?

En la actualidad la Estadística es probablemente una de las disciplinas científicas más utilizada y estudiada en todos los campos del conocimiento humano. Por ejemplo:

- en la Administración de Empresas se utiliza para evaluar la aceptación de un producto antes de comercializarlo,

- en Economía para medir la evolución de los precios mediante números índice o para estudiar los hábitos de consumo mediante encuestas,
- en Ciencias Políticas para conocer las preferencias de los electores antes de la votación mediante sondeos y así orientar las estrategias de los candidatos,
- en Sociología para estudiar las opiniones de los colectivos sociales sobre temas de actualidad,
- en Psicología para elaborar las escalas de los tests y cuantificar aspectos del comportamiento humano,
- en general en las Ciencias Sociales para medir la relación entre variables y hacer predicciones sobre ellas.

En las Ciencias Sociales la Estadística se estudia en tres secciones: la Estadística Descriptiva, la Estadística Inferencial y el Diseño Experimental. La Estadística Descriptiva sirve de herramienta para describir, resumir o reducir las propiedades de un conglomerado de datos al objeto de que se pueda manejar. La Estadística Inferencial se utiliza para estimar las propiedades de una población a partir del conocimiento de las propiedades de una muestra de ella. Y en tercer lugar, el diseño y análisis de experimentos se desarrolla para determinar y confirmar relaciones causales entre variables.

En la investigación la Estadística es importante porque:

- permite el tipo más exacto de descripción,
- fuerza a ser exactos y definidos en nuestros procedimientos y pensamiento,
- permite resumir nuestros resultados de una forma conveniente,
- permite extraer conclusiones generales,
- permite predecir, y
- permite analizar algunos de los factores causales que subyacen a eventos complejos.

Dentro del campo de la Psicología hay tres vertientes metodológicas (lo cualitativo, lo no experimental y lo longitudinal) que son el auténtico punto de partida de las actuales líneas de desarrollo. Las ecuaciones estructurales permiten la modelación de la causalidad. La regresión logística, los modelos log-lineal y el análisis de correspondencias se utilizan para el análisis de datos cualitativos. Las series temporales investigan el aspecto longitudinal.

La mayoría de las técnicas utilizadas en los cuasi-experimentos se deriva del modelo de la regresión múltiple, de modo que las hipótesis rivales son probadas una a una. Por el contrario, en los estudios aleatorizados, se estima exactamente un efecto y se eliminan otros, dado que existe garantía de que influyen por igual en el grupo experimental y en el grupo control.

### ***Estadística Descriptiva***

La estadística puede estudiar tanto las características de las muestras en sí como hacer inferencias acerca de las características de las poblaciones. Población es el conjunto de valores que tienen una propiedad común y muestra es un subconjunto (aleatorio o no) de la población. Si se estudia la variable “edad de los estudiantes que estudian cuarto de Psicología en España” la población es el conjunto de edades de todos esos estudiantes, y muestra es el conjunto de las edades de los que estudian en una de sus universidades.

Las características de las poblaciones se estudian mediante indicadores de uno o más aspectos particulares. A estos indicadores se les llama parámetros cuando se refieren a la

población, y se les llama estadísticos cuando se refieren a una muestra. Aunque en general se habla de estadísticos para referirse tanto a éstos como a los parámetros.

En la Estadística las variables se dividen en cuantitativas y cualitativas. Las variables cualitativas (o nominales) son variables de cuyos valores sólo se puede decir que son distintos. En las variables cuantitativas sus valores, además de ser distintos, se pueden ordenar (de mayor a menor). Las variables cuantitativas se dividen en tres tipos: ordinales, de intervalo y de razón. En las variables ordinales los valores son distintos y se pueden ordenar. En las variables de intervalo además de esos dos rasgos (distintos, ordenados) existe una unidad común. Y en las de razón además de esos tres rasgos existe un cero real. La afiliación política es una variable nominal. La altura o el peso son variables de razón.

En la estadística descriptiva el interés del estudio puede estar en una variable, en dos variables o en tres o más variables. En cuanto al estudio de una sola variable las características que se pueden estudiar de cada variable son:

1. Distribuciones de frecuencias a través de histogramas, gráficas de barras o polígonos de frecuencias.
2. La tendencia central, es decir, el valor más representativo de la muestra, se indica mediante la media, la mediana o la moda dependiendo del tipo de variable estudiada.
3. La dispersión de los datos, si están o no muy agrupados los números, se estudia mediante la desviación típica, la varianza, la amplitud semiintercuartílica o el rango dependiendo del tipo de variable.
4. La forma de las distribuciones, a través de los coeficientes de asimetría y apuntamiento.

La forma de la distribución toma como patrón la distribución normal o campana de Gauss. En la distribución normal la mayoría de las puntuaciones se agrupan en torno a la media, y en la que cuanto más lejos se encuentra la puntuación de la media más rara es. Además, existen igual número de puntuaciones a ambos lados de la media. Y la asimetría y el apuntamiento valen cero para una distribución normal.

Los estadísticos anteriores son aplicados a muestras, pero cuando el interés está en el individuo cada sujeto se puede estudiar con respecto a la muestra que pertenece. Se puede estudiar su puntuación directa, la puntuación diferencial (la diferencia con respecto a la media), y la puntuación típica (la distancia en desviaciones típicas a la media). La puntuación típica permite comparar verazmente las puntuaciones en distintas variables con medias y desviaciones típicas diferentes. Otro índice de la posición del sujeto en su grupo es el percentil, que indica el porcentaje de casos que tienen puntuaciones inferiores a la suya.

En el estudio de dos variables lo que interesa usualmente es hallar la relación que hay entre las variables. Se utiliza el coeficiente de correlación de Pearson cuando las variables son cuantitativas. Este coeficiente varía entre uno y menos uno; cuando vale cero indica que no hay relación entre las variables, y cuando vale uno o menos uno que la relación es perfecta. El estudio de la relación se puede complementar con su representación gráfica mediante la nube de puntos o gráfico de dispersión.

El estudio de la relación entre variables cualitativas se realiza mediante tablas de contingencia y estadísticos basados en el estadístico chi cuadrado. Hay otra serie de estadísticos cuando se intenta estudiar la relación entre variables cuantitativas y cualitativas.

Cuando se estudia la relación entre varias variables la técnica por excelencia es la regresión múltiple. Ésta da un índice, que varía entre cero y uno, de la relación entre una variable y un conjunto de variables. Otros índices estadísticos son: la correlación parcial, que estudia la relación entre dos variables cuantitativas eliminando el influjo sobre ellas de otras

variables, y el coeficiente de correlación canónica, que analiza la relación entre dos conjuntos de variables, cada uno de ellos con dos o más variables.

### ***Estadística Inferencial***

El siguiente gran capítulo de la Estadística aplicada a las Ciencias Sociales es la Estadística Inferencial. Cuando se quieren hacer inferencias, es decir, suposiciones acerca de lo que ocurre en la población global de la que se han tomado las muestras que se están estudiando, se utilizan los contrastes de hipótesis. Estos contrastes permiten contestar a preguntas como: ¿Pertenece una cierta muestra a una determinada población? ¿Pertenece a la misma población dos muestras?

Para contestar a las anteriores preguntas se adopta un nivel de significación (el máximo nivel de error que se asume en la decisión), y se compara con la probabilidad del estadístico de contraste. Entonces, si es menor la probabilidad del estadístico de contraste que el nivel de significación, se habla de que existen diferencias, o de que se han producido resultados significativos.

Este proceso se formaliza de la siguiente manera:

1. Fijar el nivel de significación o error alfa.
2. Establecer una hipótesis nula y una hipótesis alternativa.
3. Recoger los datos adecuados.
4. Hallar el estadístico de contraste.
5. Comparar la probabilidad del estadístico de contraste con el nivel de significación.

El estadístico de contraste es una cierta fórmula que da como resultado un valor. Este valor sigue una cierta distribución. Las distribuciones que siguen estos estadísticos son, en la mayoría de los casos: la distribución normal, la distribución t de Student, la distribución chi cuadrado y la distribución F.

Hay diferentes tipos de contrastes dependiendo del objetivo de estudio: diferencia entre medias, diferencia entre desviaciones típicas, diferencia entre proporciones, diferencia entre correlaciones, y así sucesivamente.

### ***Diseño Experimental***

Los principales tipos de problemas con que se enfrenta la Estadística en el campo de las Ciencias Sociales son los de: diferencia entre grupos, relación entre variables, estructura de los datos y separación entre grupos.

En el estudio de la diferencia entre grupos lo que usualmente se trata de analizar es la diferencia entre medias. Y para ello se utiliza la prueba t de diferencia entre medias o el análisis de varianza. En ambas técnicas lo que se investiga en realidad es si la diferencia entre las medias de los grupos (la llamada variabilidad intergrupos) es significativamente mayor que la diferencia dentro de los grupos (la denominada variabilidad intragrupos). El análisis de varianza ocupa una gran parte del contenido de la Estadística Inferencial en las Ciencias Sociales ya que el número y características de las variables experimentales puede ser muy variado. Y además es la base del diseño experimental. Diseño experimental que trata de establecer relaciones de causa a efecto entre variables.

En el estudio de las relaciones entre variables hay dos objetivos principales, uno de índole teórica y otro de naturaleza práctica. El primero halla índices analíticos que miden la

relación entre variables mediante las técnicas ya mencionadas: correlación de Pearson, regresión múltiple o correlación canónica. En la regresión múltiple se halla la relación entre una variable (la variable predicha o variable dependiente) y una o más variables (las variables predictoras o variables independientes). La regresión múltiple ofrece no sólo un índice numérico global de la relación, sino un estudio pormenorizado de la importancia de cada una de las variables predictoras. La cuestión práctica en el estudio de la relación entre variables consiste en poder realizar predicciones del valor de la variable predicha. Es decir, una vez establecido que existe una relación significativa y substantiva entre la variable predicha y las variables predictoras, basta con aplicar una fórmula con los valores de las variables predictoras para obtener un valor en la variable predicha sin necesidad de medirla de antemano.

El problema de la estructura de los datos se estudia básicamente con una técnica estadística multivariada: el análisis de componentes principales. Esta técnica permite reducir un conjunto de variables relacionadas a un número de componentes (variables también) independientes entre sí. Con ello se consigue agrupar las variables originales en subconjuntos de variables que están relacionadas entre sí y no están relacionadas con las variables de los otros subconjuntos. Por ejemplo, en Psicología esta técnica ha servido para constatar la existencia de distintos tipos de inteligencia: espacial, verbal, manual, etcétera.

En el problema de la separación entre grupos se aplican dos tipos de técnicas. Con la primera se trata de establecer la existencia de grupos que no son evidentes a priori; para ello se utiliza el análisis de conglomerados. Con esta técnica no sólo es posible descubrir estos grupos, sino fijar de antemano el número de ellos. Con la segunda técnica, el análisis discriminante, se estudia la estructura de la diferencia entre los grupos; puede ocurrir que la diferencia entre los grupos se explique por subconjuntos de distintas variables. Y tal como ocurre en la regresión múltiple, también se puede utilizar para predecir la pertenencia a grupos.

### **Bibliografía**

- J. Amón: *Estadística para Psicólogos: 1 Estadística Descriptiva*. Pirámide, Madrid, 1980.
- J. Arnau: Técnicas de Análisis Avanzadas y Diseño de Investigación: Tendencias Actuales y Líneas Futuras de Desarrollo. En J. Arnau (editor): *Métodos y Técnicas Avanzadas de Análisis de Datos en Ciencias del Comportamiento*. EUB, Barcelona, 1996.
- G. Glass, J. Stanley: *Métodos Estadísticos aplicados a las Ciencias Sociales*. Prentice Hall, Madrid, 1980.
- W. Hays: *Statistics for the Social Sciences*. Holt, Rinehart & Winston, London, 1973.
- D. Peña, J. Romo: *Introducción a la Estadística para las Ciencias Sociales*. McGraw-Hill, Madrid, 1997.
- A. Solanas et al.: La Enseñanza de la Estadística en las Ciencias del Comportamiento a Inicios del Siglo XXI. *Metodología de las Ciencias del Comportamiento* 4, no. 2 (2002), 157-183.