

## Análisis e interpretación de datos en investigación médica

Dr. Gerardo García Maldonado  
Doctorante en ciencias de la Salud  
Profesor de Tiempo Completo e Investigador  
Universidad Autónoma de Tamaulipas  
Facultad de Medicina de Tampico “Dr. Alberto Romo Caballero”  
[galdonado@docentes.uat.edu.mx](mailto:galdonado@docentes.uat.edu.mx)

Dr. Wilberto Sánchez Márquez  
Doctor en Desarrollo Educativo  
Profesor de Tiempo Completo e Investigador  
Universidad Autónoma de Tamaulipas  
Facultad de Medicina de Tampico “Dr. Alberto Romo Caballero”

### INTRODUCCIÓN

Hoy en día, la cuantificación es primordial para que los datos sean aplicados en forma apropiada y conveniente. En el sentir de expertos, el clínico o el investigador médico que en la actualidad no sea capaz de expresar numéricamente resultados o conclusiones, no aportará información confiable. Dicho de otra manera, la estadística como herramienta de análisis cuantitativo, en este caso, al servicio de la medicina. Podría parecer exagerada esta aseveración, pero es un hecho de realidad que la información médica que revisamos en el ámbito de nuestras competencias hoy en día, tiene el sello de medicina basada en evidencia, en donde el análisis crítico de la literatura esta enfocado en lo estadístico y metodológico para la toma de decisiones diagnósticas, pronósticas, farmacológicas, no-farmacológicas y preventivas.

David Sackett y Gordon Guyatt ambos académicos de la Universidad de Mc Master en Ontario Canadá, se percataron desde la década de 1970, de lo inasequible que era para los clínicos leer la enorme cantidad de información que se generaba, era necesario establecer estrategias de lectura científica en forma práctica, crítica y relativamente ágil. Para tal fin consideraron conveniente adiestrar a los profesionales y ofrecerles conocimientos en metodología de la investigación y bioestadística, por lo que en la década de 1990, David Sackett junto con epidemiólogos clínicos de esa universidad, inicia una serie de publicaciones de artículos-guía en la revista estadounidense JAMA, sobre la apreciación crítica de la literatura médica. Sobra decir el éxito que tuvieron sus trabajos. Vale la pena tener toda la serie de esas publicaciones. Recomendamos su búsqueda y lectura.

A través del tiempo la necesidad de practicar una lectura ágil y dinámica no solo persiste, sino que además se ha incrementado exponencialmente. La evidencia es contundente, es imperativo saber leer un artículo científico, para evaluar las mejores opciones en beneficio de los pacientes. Desde luego esto para fines médicos; pero que tal si no queremos solamente ser receptores de conocimientos, sino que aspiramos a ser generadores de nuevos conocimientos, igualmente entonces necesitaremos las competencias estadísticas y metodológicas.

Como se mencionó en el primer fascículo, se revisarán algunos procedimientos para dominar el uso apropiado de las herramientas estadísticas, así como saber y entender cuándo y cómo utilizar un determinado método, que permita interpretar la información y tomar las mejores decisiones.

En esta segunda entrega se presenta un material con tres objetivos:

- En primer término exponer una taxonomía de la estadística que sea sencilla y práctica de entender, apegándose al principio de parsimonia.
- En segundo lugar, conceptualizar y contextualizar cada uno de los elementos que integran la taxonomía. Un desglose mayor e integrativo, se irá proporcionando en fascículos subsiguientes.
- Finalmente presentar un glosario de símbolos, que es habitual encontrar en el campo de las matemáticas en general y de la estadística en particular.

### **TAXONOMÍA EN ESTADÍSTICA**

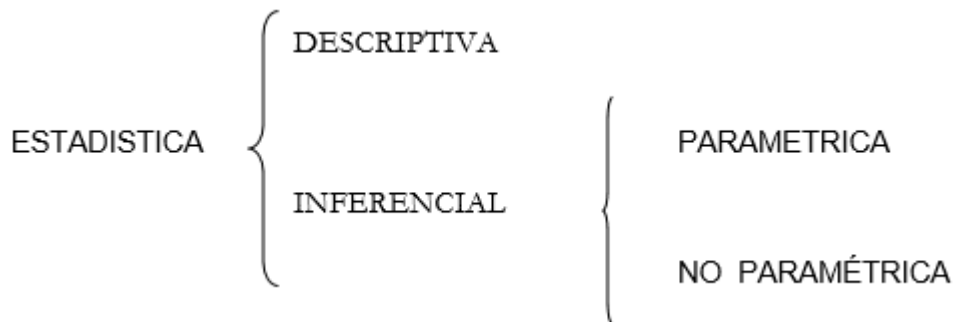
Una taxonomía es un sistema organizado de clasificación que agrupa y categoriza elementos basándose en características comunes, se considera que es sinónimo de clasificación, pero para fines de este proyecto se utilizará el vocablo taxonomía. El término proviene del griego antiguo τάξις ("ordenamiento") y νόμος ("norma o regla"). Es un procedimiento que se aplica en la ciencia en general, así como a otros campos del conocimiento. El propósito final, donde se requiera, es organizar y entender la información. Tomando en cuenta estas consideraciones la estadística no es la excepción.

Para ser efectiva una taxonomía debe ser clara, simple, relevante y compatible con normas o estándares, ya que de esta manera se asegurará que el uso de la información sea eficiente. Revisando la literatura y contra lo que pudiera pensarse, una taxonomía única en estadística no existe. Por el contrario hay diversidad de propuestas y con frecuencia se mezclan aspectos estadísticos y temática relacionada con metodología de la investigación. Por supuesto que estos elementos son complementarios en investigación científica, por lo que realmente es difícil escindirlos. Pero la idea en esta ocasión es intentarlo.

No es el objetivo primario en este fascículo proponer nuevas clasificaciones o algo atípico, simplemente se sugiere una organización que aborda exclusivamente la temática estadística. La intención en este documento fue apegarnos al principio de parsimonia, que afirma que en igualdad de condiciones la explicación más sencilla suele ser la mejor y la más conveniente.

Tomando en cuenta las características intrínsecas de los elementos señalados en esta taxonomía, se pretendió que fuera exhaustiva y excluyente, es decir, no dejar ningún elemento fuera de los grupos conformados y que dichos elementos pertenecieran sólo a un grupo.

En base a lo señalado se expone lo siguiente:



### **ESTADÍSTICA DESCRIPTIVA. CONCEPTOS Y GENERALIDADES**

La estadística descriptiva, es una pieza clave del análisis de datos. Como su nombre lo indica, su objetivo es describir y resumir las características de la información proporcionada, para que sea fácil de entender. Esta estrategia no pretende hacer inferencias ni sacar conclusiones agregadas. Este tipo de análisis implica el cálculo y la presentación de varias medidas estadísticas, como la media, la mediana, la moda, la desviación estándar, la varianza y las percentiles, que ofrecen información valiosa sobre las tendencias centrales, la dispersión y la distribución general de los datos. Otro de los objetivos principales de la estadística descriptiva, es comunicar la mayor cantidad de información de la manera más simple y eficiente posible.

Al presentar estas medidas resumidas, los lectores pueden obtener una imagen clara de los datos, incluidos los valores típicos, el grado de variabilidad y cualquier posible valor atípico o asimétrico en una muestra o una población. Su utilización se extiende a varios campos, incluida la ciencia de datos, donde juega un papel crucial en la exploración y comprensión inicial de los mismos. En el contexto de la investigación en medicina, se utilizan para

resumir características de los participantes, como la edad, raza, género, estrato social y la prevalencia de enfermedades, por solo citar unos ejemplos.

Al comprender esta información, se pueden tomar decisiones informadas para ejecutar análisis adicionales y pruebas de hipótesis, con métodos estadísticos más avanzados lo que conducirá a conclusiones más confiables.

La estadística descriptiva ofrece la ventaja de representar los datos obtenidos en gráficos de barras, histogramas, diagramas de dispersión o gráficos de líneas, entre otros métodos, lo que facilita aún más la comprensión de la información.

Podemos resumir entonces que los objetivos de esta estrategia estadística son:

- Recopilar datos
- Organizar información
- Presentar y tabular datos
- Analizar resultados

### **ESTADÍSTICA INFERENCIAL. CONCEPTOS Y GENERALIDADES**

La estadística inferencial es una rama de la estadística que permite hacer generalizaciones y tomar decisiones informadas, basadas en muestras representativas de una población en particular. La condición es que las muestras de estudio sean obtenidas de la población blanco a través de un cálculo, y que las unidades de estudio hayan sido seleccionadas en forma aleatoria. A diferencia de la estadística descriptiva, que se centra en ordenar y describir los datos recolectados, la estadística inferencial utiliza estos datos como base para realizar predicciones o inferencias, es decir, generalizaciones. Uno de los pilares fundamentales de la estadística inferencial es el concepto de probabilidad, el cual proporciona un marco para cuantificar la incertidumbre, esto es primordial ya que ningún dato es necesariamente definitivo; siempre habrá circunstancias que generen errores.

Temas como el intervalo de confianza, la estimación de parámetros y la prueba de hipótesis, son elementos clave que permitirán cuantificar resultados y hacer predicciones con mayor certeza para poblaciones completas, a partir de muestras elegidas. De cualquier manera al ser seleccionada la muestra a través de una aleatorización, cualquier conclusión estará sujeta a un margen de error, el cual debe considerarse previamente. La prueba de hipótesis, es un procedimiento central en la estadística inferencial que se utiliza para determinar si hay suficiente evidencia en una muestra de datos, para inferir que una cierta condición es verdadera para toda la población. Este proceso involucra formular una hipótesis nula y una hipótesis alternativa o de investigación, y luego usar métodos estadísticos para determinar cuál de estas hipótesis es más consistente con los datos recolectados.

Los intervalos de confianza, por otro lado, deben idealmente acompañar la cuantificación de parámetros estadísticos (medidas de tendencia central y de dispersión) y parámetros poblacionales (tasas, índices, prevalencia, incidencia, razón de momios, riesgo relativo, etc.), pues valoran el margen de error asociado al cálculo. Un intervalo de confianza del 95%, por ejemplo, sugiere que si se repitiera el muestreo muchas veces, en el 95% de los cálculos, los intervalos obtenidos contendrían el parámetro verdadero de la población. Esta metodología es crucial ya que, en el mundo real, por ser inaccesible el abordaje de datos completos de toda una población, la obtención de muestras de estudio de la población y las inferencias son más la regla que la excepción. La estadística inferencial también tiene aplicaciones en la medicina, por ejemplo, se utiliza para evaluar la efectividad de nuevos tratamientos o medicamentos.

Es crucial tener en cuenta que la estadística inferencial no está exenta de limitaciones, un muestreo inadecuado puede conducir a sesgos y en consecuencia, a inferencias incorrectas. Además, la interpretación incorrecta de los resultados estadísticos puede llevar a conclusiones erróneas y malas decisiones, en este caso: clínicas.

En resumen, la estadística inferencial es una herramienta poderosa que permite convertir datos limitados en conocimientos aplicables, para entender y mejorar el mundo que nos rodea. A través de la formulación de hipótesis, el cálculo de probabilidades y la construcción de intervalos de confianza, se proporcionará un marco robusto para la toma de decisiones en condiciones de incertidumbre.

Sin embargo, su aplicación efectiva exige una comprensión profunda de sus métodos y limitaciones para evitar errores y maximizar su potencial.

## **ESTADÍSTICA PARAMÉTRICA. CONCEPTOS Y GENERALIDADES**

La estadística paramétrica es una rama fundamental dentro de la estadística inferencial, la cual ciertamente se enfoca en el análisis de datos, pero donde estos deben cumplir ciertas reglas o supuestos (así se les denomina), para que los resultados sean confiables y válidos. Esto es básico ya que el uso de herramientas estadísticas paramétricas, garantiza un análisis más robusto y preciso de los datos en estudio. Dicho de otra manera, las pruebas estadísticas denominadas "paramétricas", se basan en un conjunto de parámetros, que se deben satisfacer incuestionablemente.

Uno de los aspectos centrales de la estadística paramétrica es el empleo de distribuciones bien conocidas, como la normal, la binomial y la de Poisson, que pueden ser representadas mediante tablas, gráficos o funciones matemáticas. En el caso de la distribución normal, que es tal vez la más conocida, debemos entender el concepto como un modelo teórico, en donde los valores de una variable o conjunto de variables en estudio, se distribuyen de manera simétrica alrededor de un valor central.

La distribución se caracteriza por una forma de campana, se describe matemáticamente y depende de dos parámetros: la media y la desviación estándar. Estas distribuciones permiten simplificar los cálculos y obtener conclusiones más rápidamente, siempre y cuando los datos se ajusten adecuadamente a los supuestos del modelo. La distribución normal es la base de numerosos métodos paramétricos, debido a su prevalencia en la naturaleza y su relación con el teorema del límite central. Este teorema establece que si se toman muchas muestras de una población, por ejemplo la talla de las personas y calculas el promedio de las mismas, a medida que el número de muestras aumenta, la distribución de esos promedios se aproximará a una curva normal, independientemente de cómo estén distribuidas dichas datos originalmente. Dicho de otra manera, cuanto mayor sea el tamaño de la muestra, más se acercará la distribución de los promedios a una distribución normal.

El uso de métodos paramétricos viene acompañado de varias ventajas; en primer lugar, cuando las reglas o supuestos para su utilización son correctos, estos métodos suelen ser muy eficientes, proporcionando estimaciones precisas y potentes con menor tamaño muestral en comparación con sus contrapartes no paramétricas. Además, debido a la estructura matemática de los modelos paramétricos, se facilitan análisis extensivos que permiten un entendimiento más profundo y riguroso de los datos.

Sin embargo, la estadística paramétrica no está exenta de limitaciones, el principal inconveniente radica en la dependencia a los supuestos sobre la distribución de los datos. Si estos no se cumplen, los resultados pueden ser incorrectos, de tal manera que es crucial verificar la distribución previo a su aplicación. Por ello, en situaciones donde no se puede garantizar que los datos sigan una distribución específica, los métodos no paramétricos pueden ser más adecuados, aunque a costa de una menor eficiencia. Entre las técnicas más comunes utilizadas para la aplicación de estadística paramétrica se incluyen las pruebas de hipótesis, la estimación de intervalos de confianza y los modelos analíticos de regresión entre otros.

Las pruebas de hipótesis, donde se pueden utilizar estadísticos como la *t* de Student y el análisis de la varianza (ANOVA por sus siglas en inglés), permiten determinar la significancia de los resultados en base a parámetros específicos, como la media o la varianza poblacional. Por su parte, los intervalos de confianza ofrecen un rango plausible para los parámetros desconocidos, proporcionando información crucial sobre su precisión y fiabilidad. Los modelos de regresión, por último, son una herramienta poderosa para modelar, entre otras cosas, la relación entre variables y hacer predicciones.

En conclusión, la estadística paramétrica constituye una herramienta esencial en el análisis de datos, dotada de métodos eficientes y poderosos para investigar y comprender fenómenos naturales y científicos. Aunque el mayor énfasis se hace en la normalidad de la distribución, existen otras consideraciones que deben tomarse en cuenta.

- Normalidad en la distribución de los datos
- Homogeneidad de varianzas (también denominada homocedasticidad)
- Independencia entre las observaciones
- Linealidad en el caso de los modelos de regresión
- Las variables deben estar medidas en una escala de intervalo o de razón.

### **ESTADÍSTICA NO PARAMÉTRICA. CONCEPTOS Y GENERALIDADES.**

La estadística no paramétrica se ha convertido en una herramienta invaluable en el análisis de datos, especialmente en situaciones donde las suposiciones tradicionales, ya señaladas, no pueden confirmarse, por lo que es más versátil y aplicable en una variedad de escenarios reales. Es importante entender qué lo que distingue a la estadística no paramétrica de su contraparte paramétrica, es que no se basa en parámetros específicos (como la media y la desviación estándar) ni asume una distribución normal de los datos. Esto es importante porque en la práctica los supuestos no siempre se pueden cumplir. Entre las principales técnicas de estadística no paramétrica se encuentran las pruebas de rango como la prueba de Wilcoxon, la prueba U de Mann-Whitney, o la prueba de H de Kruskal-Wallis. Estas pruebas comparan medianas o rangos en lugar de medias, lo que las hace más robustas en casos de datos asimétricos o con valores atípicos.

Por ejemplo, al comparar la efectividad de dos tratamientos médicos, es posible que los datos no estén normalmente distribuidos debido a la variabilidad en la respuesta de los pacientes. En tales casos, una prueba no paramétrica puede proporcionar resultados más confiables. La flexibilidad de la estadística no paramétrica, también se extiende a su capacidad para manejar diferentes tipos de datos, incluidos los datos ordinales y categóricos. A menudo en investigación social y psicológica, los datos se recogen en escalas ordinales, por ejemplo las que miden la satisfacción del paciente en una escala de "muy insatisfecho" a "muy satisfecho". Los métodos no paramétricos pueden analizar estos datos sin necesidad de convertirlos a una escala numérica, lo que preserva la integridad de la medición original.

Un aspecto crítico de las técnicas no paramétricas es su simplicidad interpretativa, es decir, al no depender de supuestos complejos sobre la distribución, los resultados pueden ser más fáciles de interpretar y comunicar a una

audiencia no técnica. Esto es particularmente beneficioso en contextos interdisciplinarios donde la claridad y la accesibilidad de los resultados son fundamentales. A pesar de sus numerosas ventajas, la estadística no paramétrica necesita muestras más grandes para alcanzar un confiable nivel de poder estadístico, además, algunos métodos no paramétricos pueden representar desafíos en casos donde hay grandes conjuntos de datos.

En conclusión, la estadística no paramétrica ofrece una alternativa poderosa y flexible a los métodos paramétricos tradicionales. Su capacidad para analizar datos sin suposiciones estrictas sobre la distribución, los hace especialmente útiles en disciplinas con datos variados y difíciles de modelar. Aunque pueden ser menos eficientes en términos de tamaño de muestra, su aplicabilidad y facilidad de interpretación, a menudo los convierten en la opción preferida en estudios reales donde las condiciones ideales rara vez se encuentran.

### GLOSARIO DE SÍMBOLOS

|   |            |
|---|------------|
| Desviación estándar muestral                            | S          |
| Desviación estándar poblacional                         | $\sigma$   |
| Varianza muestral                                       | $S^2$      |
| Varianza poblacional                                    | $\sigma^2$ |
| Media poblacional                                       | $\mu$      |
| Media muestral  | $\bar{x}$  |
| Mediana   | Md         |
| Moda  | Mo         |
| Frecuencia absoluta                                     | fai        |
| Frecuencia absoluta acumulada                           | Fai        |
| Frecuencia relativa                                     | fri        |
| Frecuencia relativa acumulada                           | Fri        |
| Frecuencia de un puntaje o categoría                    | f          |
| Grados de libertad                                      | gl         |
| Hipótesis alterna                                       | H1         |
| Hipótesis nula  | Ho         |
| Nivel de significancia                                  | $\alpha$   |
| Tamaño de la población                                  | N          |
| Tamaño de una muestra                                   | n          |
| Operador sumatorio                                      | $\Sigma$   |
| Probabilidad de cometer error tipo II                   | $\beta$    |
| Probabilidad de cometer error tipo I                    | $\alpha$   |
| Probabilidad, representa también nivel de significancia | $\rho$     |
| Proporción  | P          |
| Análisis de varianza (ANOVA)                            | F          |
| Puntuación z  | z          |
| Chi cuadrada  | $X^2$      |

|  |            |
|--|------------|
| Coefficiente de correlación de Spearman        | $\rho$     |
| Coefficiente de correlación de Pearson         | $r$        |
| Prueba de Kruskal- Wallis                      | H          |
| Prueba de T de rangos asignados de Wilcoxon    | T          |
| Prueba Q de Cochran                            | Q          |
| Prueba t de student                            | t          |
| Prueba U de Mann- Whitney                      | U          |
| Proporción                                     | P          |
| Coefficiente de Cramer                         | C          |
| Estadístico Kappa                              | K          |
| Valor i-esimo de una variable                  | $X_i$      |
| Valor máximo de una variable                   | $X_{max}$  |
| Valor mínimo de una variable                   | $X_{min}$  |
| Variable para calcular                         | x          |
| Límite superior de un intervalo                | $L_s$      |
| Límite inferior de un intervalo                | $L_i$      |
| Amplitud de un intervalo                       | a          |
| Marca de clase                                 | X          |
| Error aleatorio o residual en regresión lineal | $\epsilon$ |

## REFERENCIAS

- A BRIEF HISTORY OF STATISTICS (SELECTED TOPICS). ALPHA Seminar. 2017, 29 Ago [citado 08/07/2024]. Disponible en: <http://homepage.divms.uiowa.edu/~dzimmer/alphaseminar/Statistics-history.pdf>
- Barrios JA. KARL PEARSON, CREADOR DE LA BIOMETRIA, PRECURSOR DE LA ESTADISTICA CONTEMPORANEA. Cizur Menor, España: Editorial Aranzadi; 2023.
- Betanzos FG, de León M del CEP, López JKC. ESTADÍSTICA APLICADA EN PSICOLOGÍA Y CIENCIAS DE LA SALUD. Hipodromo: El Manual Moderno, Editorial; 2017.
- Daniel WW. BIOESTADISTICA: BASE PARA EL ANALISIS DE LAS CIENCIAS DE LA SALUD. Editorial Limusa S.A. De C.V.; 2002.
- De la Fuente Solana Joan Guàrdia Olmos EI. HISTORIA DE LA ESTADÍSTICA: UN EJERCICIO PARA ENTENDER EL MUNDO ACTUAL. España: Universidad de Barcelona; 2021.
- De Mora Charles MAGVMS. HISTORIA DE LA PROBABILIDAD Y LA ESTADÍSTICA. España: UNED - Universidad Nacional de Educación a Distancia; 2018.
- Del Cerro JS. HISTORIA DE LA PROBABILIDAD Y DE LA ESTADÍSTICA XII. Cizur Menor, España: Editorial Aranzadi; 2023.

- Fernández JS-L. HISTORIA DE LA ESTADÍSTICA COMO CIENCIA EN ESPAÑA, (1500-1900). España: Instituto Nacional de Estadística; 1975.
- Koren J. THE HISTORY OF STATISTICS, THEIR DEVELOPMENT AND PROGRESS IN MANY COUNTRIES: IN MEMOIRS TO COMMEMORATE THE SEVENTY-FIFTH ANNIVERSARY OF THE AMERICAN STATISTICAL ASSOCIATION. USA: American Statistical Association; 1918.
- Lopez FJB. BIOESTADÍSTICA. España: Ediciones Paraninfo, S.A; 2005.
- Macho JMA. HISTORIA DE LA ESTADISTICA. España: UNED - Universidad Nacional de Educación a Distancia; 2006
- Martín MA, Horna O, Nedel F, Navarro A. FUNDAMENTOS DE ESTADISTICA EN CIENCIAS DE LA SALUD. Barcelona: Servei de Publicacions de la Universitat Autònoma de Barcelona; 2010.
- Miranda RHC. BIOESTADÍSTICA APLICADA A LAS CIENCIAS DE LA SALUD. España: Formación Alcalá, S.L.; 2017.
- Stigler SM. THE HISTORY OF STATISTICS: THE MEASUREMENT OF UNCERTAINTY BEFORE 1900. Cambridge, Mass., Estados Unidos de América: Belknap Press; 1990.
- Vilalta JS. MANUAL DE BIOESTADÍSTICA. Barcelona, Spain: Elsevier Masson; 2003.