

ENRIQUE VILLARREAL RÍOS

# BIOESTADÍSTICA APLICADA AL DISEÑO DE INVESTIGACIÓN





ENRIQUE VILLARREAL RÍOS

# BIOESTADÍSTICA APLICADA AL DISEÑO DE INVESTIGACIÓN



Madrid • Buenos Aires • México • Bogotá

©Enrique Villarreal Ríos, 2026 (edición ebook)

Reservados todos los derechos.

«No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright.»

Ediciones Díaz de Santos

Internet: <http://www.editdiazdesantos.com>

E-mail: [ediciones@editdiazdesantos.com](mailto:ediciones@editdiazdesantos.com)

ISBN: 978-84-9052-596-8 (edición papel)

e-ISBN: 978-84-9052-597-5 (edición digital)

Depósito Legal: M-12557-2026

Diseño de cubierta y Fotocomposición: P55 Servicios Culturales

A Melissa

*A dónde miras paloma  
Que no me quieres mirar  
Tú tienes las golondrinas  
Yo tengo tu libertad  
Mírame Niña Paloma  
Mira mis ojos  
Aquí tenemos la mar*

JUAN ANTONIO VILLARREAL



## AUTOR

---

- Licenciado en Medicina, Universidad Autónoma Metropolitana
- Especialista en Medicina Familiar, Universidad Nacional Autónoma de México
- Maestro en Ciencias en Sistemas de Salud, Instituto Nacional de Salud Pública
- Maestro en Filosofía, Facultad de Filosofía, Universidad Autónoma de Querétaro
- Director, Unidad de Investigación Epidemiológica y en Servicios de Salud Querétaro, Instituto Mexicano del Seguro Social
- Profesor de Bioestadística, Residencia de Medicina Familiar Querétaro, Instituto Mexicano del Seguro Social Querétaro
- Profesor de Bioestadística, Maestría en Química Clínica, Facultad de Química, Universidad Autónoma de Querétaro
- Miembro de la Comisión de Investigación Científica del Colegio de Médicos Familiares y Generales de Querétaro

Datos de localización

Correo electrónico: [enriquevillarrealrios@gmail.com](mailto:enriquevillarrealrios@gmail.com)



## PROEMIO

.....

En la sociedad hipermoderna el discurso incluye el mercado, los derechos del hombre, el individualismo y la tecnociencia, sin embargo en ella lo que importa es el fin, la forma para llegar al fin no interesa, es relegada a segundo término.

La Investigación Científica no escapa a esta tendencia; en particular en el caso de la bioestadística las herramientas electrónicas, entiéndase por ello programas estadísticos de cómputo, son un claro ejemplo. Su uso permite ahorrar tiempo, procedimientos que anteriormente requerían de días o en ocasiones semanas, en la actualidad se pueden realizar en minutos. Esto, evidentemente, es un gran avance.

No obstante, la tecnificación excesiva, particularmente en el terreno académico, conceptual y de la razón, se acompaña de riesgos. Olvidar o no entender el marco conceptual, olvidarse del aspecto epistemológico, olvidarse de la lógica que da sustento al proceso técnico es un grave riesgo, es quedar vacío y existir en la nada.

En la actualidad, con los avances tecnológicos cualquier persona que se apoye en un programa estadístico puede realizar la prueba estadística que desee, pero esto es solamente técnica. Para ello no se requiere de un marco conceptual, ni de un proceso lógico, basta con oprimir una tecla y el resultado estará disponible. Eso no es realizar análisis estadístico. Los programas estadísticos son solamente una herramienta que de ninguna manera sustituyen al uso de la razón, no obstante, de pronto pareciera que esta, la razón, es desplazada por la herramienta tecnológica y entonces la razón pasa a un segundo plano, cuando no a desaparecer.

En estos tiempos, en el análisis estadístico que se hace a los proyectos de investigación pareciera que el referente mágico es la significancia estadística, el valor 0.050. Identificarlo permite aceptar o rechazar la hipótesis nula, pero no necesariamente significa entender la lógica que respalda la decisión. Este panorama convierte al investigador en un individuo que replica conceptos sin entender el significado, los reproduce por tradición, lo cual es muy grave ya que lo aleja del uso de la razón en sentido extenso.

En este texto, cuando se plantea entrar al mundo de la bioestadística no se propone como una actividad aislada, alejada del mundo, alejada de otras actividades que cotidianamente se presentan, como si se tratara de un ermitaño en el terreno académico. Tampoco se propone abordarla como una actividad vacía, fría, sin sustento conceptual y lógico.

La propuesta contempla abordar la bioestadística con relación al todo (aunque el todo es temporal, es relativo, es cambiante y probablemente el todo no existe), desde la perspectiva de la teoría general de sistemas, como una de las partes que integran el todo y sin la cual no se puede llegar al todo. En este caso, en el terreno académico, cuando se habla del todo se está hablando del conocimiento.

De igual forma, en relación a ella misma, a la bioestadística misma como una actividad que obedece a un marco conceptual y a una lógica que le permite alcanzar el fin, lo que se pretende destacar son los medios que emplea, más allá del fin; y aún más, la propuesta del libro no es abordar la bioestadística como bioestadístico, la propuesta pretende llevar la bioestadística del mundo teórico al mundo aplicado, propuesta que significa alejarse de algunos preceptos que para el bioestadístico son incuestionables pero que al momento de convertir la bioestadística en una parte del proceso de investigación científica, la perspectiva puramente teórica se deja atrás y se aterriza en la bioestadística aplicada.

Como una parte del todo la bioestadística se presenta íntimamente relacionada con la pregunta de investigación, el objetivo, la hipótesis y el diseño empleado. La elección de la prueba estadística necesariamente se debe dar en esta lógica, pregunta-objetivo-hipótesis-diseño. El investigador está obligado a reconocerla, debe hacer consciente la relación que precede a la prueba estadística, en el contexto de la investigación; la elección de la prueba estadística no se puede realizar fuera de este contexto.

Por otro lado, al interior de la prueba estadística también es necesario identificar el marco conceptual y la lógica que respaldan a la prueba estadística. Es decir, la relación información, tipo de distribución, hipótesis estadística, teórica, hipótesis estadística aplicada, hipótesis de investigación, prueba estadística, criterios de referencia, cálculo de la prueba estadística, decisión estadística, nivel de confianza, valor de  $p$ , tipo de error, conclusión de investigación.

En este sentido la propuesta del libro no pretende el resultado de la prueba estadística, la propuesta va en pos de la aprehensión de la lógica para la elección de la prueba estadística en el contexto de la investigación

científica; y al interior de la prueba, la propuesta va en pos de la aprehensión de la lógica empleada, desde el planteamiento pasando por la decisión estadística para llegar a la conclusión de investigación, un concepto. En este contexto, en cada una de las pruebas estadísticas presentadas se replican los pasos, pretendiendo así alcanzar la sistematización sin alejarse de la razón.

**Enrique Villarreal Ríos**



# ÍNDICE

|                                                                         |            |
|-------------------------------------------------------------------------|------------|
| <b>Autor .....</b>                                                      | <b>VII</b> |
| <b>Dedicatoria .....</b>                                                | <b>IX</b>  |
| <b>Proemio.....</b>                                                     | <b>XI</b>  |
| <br>                                                                    |            |
| 1. Generación del conocimiento .....                                    | 1          |
| 2. Variables.....                                                       | 11         |
| 3. Técnicas de conteo .....                                             | 15         |
| 4. Distribución binomial.....                                           | 19         |
| 5. Medidas de tendencia central y medidas de dispersión .....           | 23         |
| 6. Distribución normal .....                                            | 33         |
| 7. Tamaño de la muestra .....                                           | 41         |
| 8. Análisis estadístico en investigación.....                           | 57         |
| 9. Elección de la prueba estadística inferencial .....                  | 63         |
| 10. Diseños de investigación.....                                       | 69         |
| 11. Prueba de hipótesis .....                                           | 75         |
| 12. Chi cuadrada .....                                                  | 85         |
| 13. Razón de momios e intervalo de confianza de razón<br>de momios..... | 93         |
| 14. Riesgo relativo e intervalo de confianza de riesgo relativo .....   | 103        |
| 15. Intervalos de confianza para porcentajes y promedios .....          | 111        |
| 16. Prueba de t de student .....                                        | 123        |
| 17. Prueba de Z para prevalencias .....                                 | 139        |
| 18. Prueba de análisis de varianza (para muestras independientes).....  | 147        |
| 19. Prueba de Mann-Whitney .....                                        | 159        |
| 20. Prueba de Wilcoxon para muestras pareadas (antes-después).....      | 165        |
| 21. Prueba de Kruskal-Wallis.....                                       | 173        |
| 22. Prueba de Friedman.....                                             | 181        |
| <br>                                                                    |            |
| <b>Bibliografía .....</b>                                               | <b>189</b> |
| <b>Anexos .....</b>                                                     | <b>193</b> |



## GENERACIÓN DEL CONOCIMIENTO

---

Hablar del mundo circundante, hablar de la realidad, es hablar de una estructura involucrada en un proceso permanente de cambio influenciado por la multiplicidad de factores en él contenidos, el ser humano incluido entre ellos. Este personaje, a diferencia del resto de los factores y del resto de los seres vivos, en un momento del devenir decide alejarse del mito y del dogma para tener una explicación racional y objetiva de la realidad. En ese momento inicia un proceso de búsqueda para entender, explicar, interpretar y conceptualizar la realidad, el mundo, todo lo en él contenido. En ese momento busca generar conocimiento.

La realidad, con todas las relaciones incluidas en ella, existe desde siempre y existirá más allá de ser accesible para el humano, y aún más, cambiará con el tiempo y seguirá siendo realidad que en algún momento del tiempo se encontrará con el ser humano para ser explicada, interpretada y conceptualizada. Pero aun en ese momento la realidad continuará siendo la realidad, más allá de la correcta o incorrecta interpretación ofrecida por el ser humano. Se podría decir que es una verdad temporal y por lo tanto cambiante.

La forma de llegar al conocimiento es diversa. Al respecto las posturas filosóficas, entre otras, van del empirismo, al racionalismo, al positivismo, al criticismo y al realismo; posturas cada una con un enfoque diferente pero con argumentos lógicos que las mantienen vigentes. Algunas posturas privilegian la evidencia, otras la razón, otras más la experiencia, otras la relación entre experiencia y razón, y aquellas que privilegian al objeto. Posturas todas ellas que teorizan y analizan el método empleado para llegar al conocimiento en la práctica cotidiana. Al respecto, el método científico y la investigación científica es una forma.

La razón de ser de la investigación científica empleando el método científico es generar conocimiento, identificar la verdad, propuesta que equivale a identificar la relación existente entre los factores; en el sentido

más puro es establecer la relación entre causa y efecto, es decir identificar la causalidad. Pero en el sentido amplio de generar conocimiento y establecer la verdad, la investigación científica no se limita a la relación de causalidad. En este contexto, puede alcanzar la asociación entre factores sin que se determine el sentido de la relación, o puede alcanzar la diferencia de comportamiento entre los factores de estudios, e incluso puede centrarse en uno solo de los factores y describir las características del propio factor y de su entorno. Desde esta perspectiva la generación del conocimiento rebasa los límites de la causalidad y confronta el concepto clásico del término en el ámbito científico; es verdad que no identifica la relación causal, identifica la relación entre variables o identifica el comportamiento de una sola variable.

En torno al conocimiento se debe identificar dos perspectivas, la generación del conocimiento y la adquisición del conocimiento.

En el primero, en la generación de conocimiento, el ser humano es simple espectador a pesar de estar activamente involucrado en el proceso de generación. En ese caso el humano no importa, lo importante es el comportamiento o la relación que se establece entre los factores que integran la realidad, condición que existe antes de ser identificada por el ser humano; esto significa que la realidad o el conocimiento va más allá del humano y antes del humano.

Después de la generación del conocimiento, objetivo de la ciencia, la adquisición de ese conocimiento por el ser humano sigue otro proceso totalmente diferente para incorporarlo al consciente personal del ser humano como individuo o para incorporarlo al consciente colectivo de la sociedad. Este proceso es independiente del proceso científico, no obstante puede ser objeto de estudio.

La línea entre generación de conocimiento y adquisición del conocimiento parece ser muy delgada y en ocasiones superponerse, no obstante el científico y el pedagogo tienen clara la diferencia, y distinguen el método científico y el método pedagógico. En el primero se genera el conocimiento y en el segundo se incorpora el conocimiento generado, y si se continuara con el proceso, este se extiende a la utilización del conocimiento que en ese momento ya se ha transformado en información. Ese es otro tema.

En la literatura se han descrito características indispensables para definir cuándo el conocimiento es conocimiento. En torno al abordaje para la explicación del mundo natural o social existen dos propuestas.

La primera propuesta postula el papel del Investigador desde una perspectiva activa física (experimentación), en ella el investigador modifica deliberadamente las condiciones y comportamiento de una de las variables que intervienen en el fenómeno de estudio, actividad siempre complementada por la observación y el análisis racional. Evidentemente la propuesta implica tomar una parte de la realidad y modificarla en un punto específico, pero siempre sustentada en un marco conceptual. En este caso la realidad se modela con características propias, aunque en ocasiones ese modelaje de la realidad es perspectiva del investigador sin que necesariamente sea compartida por el resto, pero invariablemente de ello plantea una visión a futuro.

En la segunda propuesta el papel del investigador se centra en la observación y el análisis racional. En ella el investigador no modifica intencionalmente el comportamiento de las variables que determinan el fenómeno; la experimentación queda fuera de este proceso. Claro que se podría argumentar que antes del análisis racional existe un proceso de experimentación natural en el que las variables interactúan de una determinada manera, es verdad la aseveración, por lo tanto es obligación del investigador identificar ese proceso para poder establecer el comportamiento de la relación causal existente entre las variables.

Si bien en lo señalado la relación de las variables identifica la causa y el efecto para ubicarse en el contexto de la causalidad, cuando se genera conocimiento basado en la observación la relación puede establecerse en una relación de asociación no causal, pero a fin de cuentas relación entre variables; o puede establecerse en términos de diferencia entre variables o grupos, lo cual también es conocimiento; y finalmente, en el contexto de la investigación observacional, el conocimiento se puede representar por la identificación del comportamiento de un fenómeno en particular.

En ambas propuestas el producto siempre es la generación de conocimiento en cualquiera de sus dimensiones, la causalidad, la asociación no causal, la diferencia o la prevalencia.

Ante este panorama y en el contexto de un mundo probable y cambiante, donde la verdad es efímera en el largo plazo, cada una de las dimensiones que se pueden alcanzar en el proceso de generación del conocimiento tienen características propias, algunas de ellas indispensables para asegurar qué es conocimiento; o tal vez de lo que se debería hablar es de la verdad temporal en cualquiera de las cuatro dimensiones señaladas. En consecuencia, se deben identificar los criterios que necesariamente

deben estar presentes para hablar de verdad temporal. Esta propuesta va más allá de los criterios de causalidad que en la literatura erróneamente se han adoptado como sinónimo de conocimiento.

Existen criterios que son imprescindibles para poder asegurar que el conocimiento es conocimiento; existen criterios generales que son imprescindibles y que deben estar presentes en cualquiera de las dimensiones: criterios propios de la prevalencia, criterios propios de la diferencia, criterios propios de la asociación no causal, criterios propios de la causalidad y criterios propios de la experimentación.

## CRITERIOS IMPRESCINDIBLES PARA HABLAR DE CONOCIMIENTO

- **Plausibilidad.** En el contexto científico la plausibilidad se refiere a la explicación funcional específica de una relación o de un evento. En el ámbito clínico corresponde al sustento fisiopatológico, si el ámbito es el económico el sustento necesariamente tendrá que ser económico, y así respectivamente para cada disciplina. La plausibilidad es indispensable en el proceso de generación de conocimiento. Desde esta perspectiva la plausibilidad debe estar presente en la causalidad, en la asociación no causal, en la diferencia y en la prevalencia. Su ausencia elimina la posibilidad de hablar de conocimiento.
- **Significancia estadística.** Si la explicación se realiza desde la perspectiva estadística, entonces es indispensable que se encuentre presente la significancia estadística, aquella relación que cuente con sustento teórico pero que no se logre demostrar estadísticamente; desde esta perspectiva, no se puede considerar como conocimiento. Al respecto el punto de referencia adoptado generalmente se establece en 95% pero este referente podrá ser modificado. La significancia estadística debe estar presente en la causalidad, en la asociación no causal, en la diferencia y en la prevalencia.
- **Coherencia con el conocimiento.** Hablar de consistencia con el conocimiento existente significa que el resultado de la investigación coincide con los resultados aceptados como verdad por la comunidad científica, esto sin duda es inobjetable. No obstante, si una de las características de la ciencia postula que el conocimiento es falible y temporal, entonces el resultado de una nueva investigación puede no ser coherente con el conocimiento existente y se puede llegar a convertir en nuevo conocimiento que substituirá al viejo conocimiento;

para que esta sustitución se materialice, el resultado de una nueva investigación que diverja de la tendencia predominante en ese momento del tiempo, deberá ser reforzado por estudios con resultados similares. En ese momento el nuevo conocimiento sustituirá al viejo conocimiento y con ello el cambio de paradigma, planteamiento que involucra todas las dimensiones del conocimiento: causalidad, asociación no causal, diferencia y prevalencia.

- **Divergencia con el conocimiento existente.** Derivado de lo señalado en la coherencia del conocimiento, por contradictorio que pueda parecer, la generación del conocimiento requiere en determinado momento romper con lo aceptado como verdad y proponer una nueva verdad, la cual invariablemente deberá estar respaldada por resultados en el mismo sentido que vayan fortaleciendo y constituyendo ese nuevo concepto como verdad. Este planteamiento está presente en todas las dimensiones del conocimiento: causalidad, asociación no causal, diferencia y prevalencia.

## CRITERIOS PROPIOS DE LA PREVALENCIA

- **Prevalencia.** Es la ocurrencia de un fenómeno determinado en un evento determinado y en un momento determinado. Se estima para un solo grupo; al interior de ese grupo no se compara, pero sí es comparable con la ocurrencia del fenómeno publicado en otros grupos o poblaciones. Están presentes únicamente en el diseño transversal descriptivo.

## CRITERIOS PROPIOS DE LA ASOCIACIÓN CAUSAL O NO CAUSAL

- **Fuerza de asociación entre variables.** La fuerza de asociación corresponde a la intensidad de la relación que se establece entre las variables; cuanto más alta, mayor la fuerza. En el diseño experimental aleatorizado, el cuasi experimental y la cohorte, se identifica con el riesgo relativo y está determinada por la relación entre los expuestos enfermos y los no expuestos enfermos; cuanto más grande sea la diferencia entre los grupos de comparación, la fuerza de asociación será mayor. En estos tres diseños la fuerza de asociación es causal. En el diseño de casos y controles la fuerza de la asociación también es causal y se identifica por la razón de momios, que es una aproxi-

mación al riesgo relativo y está medida a partir de los enfermos (caso) expuestos y los no enfermos (control) expuestos.

En el diseño transversal analítico la fuerza de asociación también está presente, pero a diferencia de los diseños anteriores la fuerza de la asociación es no causal y se representa mediante la razón de momios. En el diseño transversal descriptivo, caracterizados por poseer un solo grupo, no se puede medir fuerza de asociación.

- **Especificidad de la asociación entre variables.** En términos estrictos, la especificidad de la asociación identifica una exposición específica y la consecuencia específica derivada de la exposición. Al respecto se deberá reconocer que en un ambiente controlado como es el laboratorio, la especificidad es clara y directa, la presencia de la causa, significa obligadamente la presencia del efecto. No obstante, en la investigación clínica, epidemiológica y en servicios de salud, al estudiar grupos de personas, poblaciones o sistemas, en las que no se tienen controladas el resto de las características, se incursiona en el terreno de los síndromes, en consecuencia la especificidad se traslada al terreno de la probabilidad, lo que significa que ante la presencia del factor (causa) existe determinada probabilidad de ocurrencia del efecto. En buena medida dependerá del propio factor pero no se puede negar que, en compañía de otras variables, esa misma característica puede no ser específica en la relación.

Evidentemente lo señalado cuestiona el concepto de causa suficiente, no así el de causa necesaria. Es verdad que aquí se ha planteado como relación causa efecto, pero en escenarios de no causalidad, caracterizados por no identificar la direccionalidad de la relación, como sucede en los diseños de asociación no causal o de diferencia, la explicación ofrecida anteriormente no tiene cabida.

Sin duda en el diseño transversal descriptivo no está presente la especificidad de la asociación entre variables por corresponder a una sola variable.

- **Consistencia de la asociación.** La consistencia de la asociación se alcanza cuando diferentes estudios realizados en torno al tema, presenta resultados similares. Al respecto, deberá señalarse que esta característica no es exclusiva de los diseños de causalidad, para fines prácticos es un requisito de cualquier diseño de investigación, causal, asociación o diferencia. Plantear la consistencia de la asociación como requisito indispensable para la generación del conocimiento tiene implicaciones:

la afirmación pone en duda la posibilidad de generar nuevo conocimiento. Si el producto de la investigación es totalmente diferente a lo previamente conocido, bajo esta afirmación se descartaría, y en ese escenario nunca existiría la posibilidad de generar nuevo conocimiento. También es verdad que un solo resultado de investigación no necesariamente se puede constituir como verdad, se requiere el mismo resultado en varias investigaciones. Este es un tema para debatir.

- **Gradiente.** El gradiente involucra el grado de exposición y la consecuencia. Así planteado, el gradiente tiene explicación en el ensayo aleatorizado, cuasi experimental, cohorte y casos y controles, diseños en los que se alcanza la causalidad; pero en un sentido extenso el gradiente rebasa la dimensión de causalidad y tiene cabida en el diseño transversal analítico, el problema en este caso se presenta por la ausencia de una causa y un efecto. En consecuencia, es difícil definir a quién se le asigna la exposición y a quién la respuesta. En el diseño transversal descriptivo, en el contexto de la relación entre dos variables, el gradiente no se puede medir.

## CRITERIOS PROPIOS DE LA CAUSALIDAD

- **Temporalidad.** Hablar de temporalidad significa que la variable independiente (causa) ocurrió antes de la ocurrencia de la variable dependiente (efecto). Esta relación se presenta de manera natural en el diseño aleatorizado, cuasi experimento y cohorte. En el diseño de casos y controles también está presente, pero en ellos el investigador debe asegurar la temporalidad; si no lo logra, entonces el diseño será transversal analítico y no existirá causalidad. Además de este diseño, el diseño transversal descriptivo tampoco alcanza temporalidad.

## CRITERIOS PROPIOS DE LA INTERVENCIÓN

- **Experimentación.** Experimentación implica la manipulación de la variable independiente (causa) por parte del investigador para medir la consecuencia en la variable dependiente (efecto), escenario que se limita al ensayo aleatorizado y cuasi experimento.

Ante este panorama, en el proceso de generación del conocimiento se puede hablar de verdad temporal en cualquiera de las dimensiones: causalidad, asociación no causal, diferencia o prevalencia.

**Tabla 1.1.** Tipo de dimensión alcanzada y diseños de investigación epidemiológica.

| Conocimiento                      | Dimensión               | Abordaje      | Diseño              |                         |
|-----------------------------------|-------------------------|---------------|---------------------|-------------------------|
| Conocimiento<br>(verdad temporal) | Causalidad              | Experimental  | Ensayo aleatorizado |                         |
|                                   |                         |               | Cuasi experimento   |                         |
|                                   | Asociación o diferencia | Observacional | Analítico           | Cohorte                 |
|                                   |                         |               |                     | Casos y controles       |
|                                   | Prevalencia             |               |                     | Transversal analítico   |
|                                   |                         |               | Descriptivo         | Transversal descriptivo |