



sincie

Sociedad Hispana de
Investigadores Científicos



NIVELES DE INVESTIGACIÓN

Una clasificación emergente fundamentada en el análisis
de datos para la investigación científica

José Supo

Sociedad Hispana de Investigadores Científicos

2025

NIVELES DE INVESTIGACIÓN

Una clasificación emergente fundamentada en el
análisis de datos para la investigación científica

José Supo

Sociedad Hispana de Investigadores Científicos

BIOESTADISTICO

2025

CRÉDITOS

Título de la publicación:

Niveles de investigación: Una clasificación emergente fundamentada en el análisis de datos para la investigación científica

Número de edición:

Primera edición

Autor principal:

José Antonio Supo Condori

Editorial / Sello:

BIOESTADISTICO

Razón social y domicilio legal del editor e impresor:

BIOESTADISTICO Empresa Educativa EIRL, Av. Sánchez Carrión 615, Oficina 1008, Jesús María, Lima, Perú.

Mes y año de edición:

Julio de 2025

Tiraje de la publicación (edición Perú):

500 ejemplares

Depósito Legal:

Biblioteca Nacional del Perú N° 2025-09491

ISBN:

978-612-99163-0-9

Derechos y licencia

© BIOESTADISTICO, 2025. Esta obra está bajo una licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0). Usted es libre de compartir este libro, copiarlo y redistribuirlo en cualquier medio o formato, siempre que: atribuya correctamente la autoría, no haga uso comercial y no genere obras derivadas.

Más información en: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Cómo citar este libro

APA 7^a edición:

Supo, J. A. (2025). Niveles de investigación: Una clasificación emergente fundamentada en el análisis de datos para la investigación científica (1.^a ed.). BIOESTADISTICO.

Vancouver:

Supo JA. Niveles de investigación: Una clasificación emergente fundamentada en el análisis de datos para la investigación científica. 1^a ed. Lima: BIOESTADISTICO; 2025.

Nota de impresión

Edición Perú (imprensa local): Impreso en el Perú

Edición internacional (Amazon KDP): Printed in the USA by Amazon KDP

RED INTERNACIONAL DE COAUTORES

Este libro es resultado de un esfuerzo colectivo. Investigadores y profesionales de distintos países participaron como coautores, aportando su conocimiento y experiencia en el análisis de datos y la metodología científica. Gracias a esta colaboración internacional, la clasificación de niveles de investigación adquiere legitimidad y relevancia práctica en múltiples contextos.

Guillermo Ramírez Hernández	Universidad Para el Bienestar Benito Juárez García	México
Nayely Reyes Noriega	Hospital Infantil de México Federico Gómez	México
María Elizabeth Rábago Sánchez	Universidad Juárez del Estado de Durango	México
Jorge Flores Barbosa	Universidad ION	México
Elia Fabiola Hernández Sánchez	Universidad Juárez del Estado de Durango	México
Cynthia Paola López Murillo	Universidad Juárez del Estado de Durango	México
Walfred Rueda Medina	Universidad Nexum	México
Laura Helena Peyro Arenas	Hospital General de Durango - Secretaría de salud	México
Gabriela Tepox Romero	Instituto Mexicano del Seguro Social	México
Olga Lidia Valenzuela Limón	Universidad Veracruzana	México
Pedro Alberto Mendoza Pablo	Universidad del Ejército y Fuerza Aérea	México
Rosa Elvira Minchala Urgiles	Universidad Católica de Cuenca	Ecuador
Gloria Ruiz Guzmán	Universidad Autónoma Metropolitana	México
Marcos Antonio Medina Pulido	Hospital Integral de Kantunilkin	México
Teodocio Gaitán Ruiz	Universidad Autónoma de Chiriquí - Centro de Investigación de Ciencias Médicas y Bioquímicas	Panamá
Jordi Omar Cárdenas Vargas	Universidad Nacional Autónoma de México - Facultad de Estudios Superiores Iztacala	México

Erick Dionicio Araúz Ortiz	Universidad Autónoma de Chiriquí	Panamá
José Antonio Díaz Velázquez	Universidad Autónoma Metropolitana - Unidad Xochimilco	México
Alma Guadalupe Arellano Meneses	Universidad Autónoma Metropolitana	México
Carolina María Antonieta Romo González	Instituto Nacional de Pediatría	México
María Guadalupe Casales Hernández	Universidad Autónoma Metropolitana - Unidad Xochimilco	México
Gladys Inés Bustamante Cabrera	Universidad Mayor de San Andrés	Bolivia
Gonzalo Mansilla Canelas	El Hospital Materno Infantil de la Caja Nacional de Salud	Bolivia
María Martina Kantún Can	Senado de la República Mexicana	México
Carmen Gorety Soria Rodríguez	Hospital General de Mexicali / IMSS Bienestar de Baja California.	México
Javier Alonso Trujillo	Universidad Nacional Autónoma de México - Facultad de Estudios Superiores Iztacala	México
María Elena Haro Acosta	Universidad Autónoma de Baja California	México
Gladys Esther Patiño Villalva	Universidad de Cuenca	Ecuador
Berlín Gricel Cáceres Rodríguez	Universidad Nacional Autónoma de Honduras	Honduras
Rubí Estela Morales Salas	Universidad de Guadalajara	México
Jesús Martínez Ángeles	Instituto Mexicano del Seguro Social. OOAD Hidalgo	México
Erick Calva Castañeda	Universidad Autónoma del Estado de Hidalgo	México
Nora Asunción Aquino González	Instituto IDEIA	Paraguay
Ana Guillermina Navarro Almeida	QSS Consulting	México
Valdemar Tapia Núñez	QSS Consulting	México
Jorge Ceferino Carrizo	Instituto Universitario de Ciencias de la Salud Fundación Barceló - Facultad de Medicina	Argentina

Bernardo Arteaga de la Garza	Investigador independiente	México
Laura Hermila De la Garza Salinas	Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado de Nuevo León	México
Abraham Alonso Ricardez	Universidad Nacional Autónoma de México - Facultad de Estudios Superiores Iztacala	México
Arabi Eduardo Soriano González	ICONOS - Instituto de Investigación en Comunicación y Cultura	México
Carmen Otilia Rodríguez Quiel	Universidad Especializada de las Américas	Panamá
Celia Emma Solís Guzmán	Universidad Nacional Autónoma de México	México
Martha Leticia Rujano Silva	Universidad de Guadalajara	México
Claudia Delfín Ruiz	Centro Universitario del Sur de la Universidad de Guadalajara	México
Porfirio Visoso Palacios	Hospital Central Sur de Alta Especialidad	México
Norman Antonio Flores García	Universidad Evangélica de El Salvador	El Salvador
Martha Lilia Martínez Servín	Hospital de Petróleos Mexicanos Central Sur	México
Alberto Méndez Martínez	Universidad La Salle	México
Eréndira González Orozco	Hospital General ISSSTE / Tepic Nayarit	México
Ernestina Amanda López Cruz	Instituto Mexicano del Seguro Social	México
Alma Rosa Márquez Fiol	Hospital General de Mexicali / IMSS Bienestar, Baja California	México
Emanuel Gómez Martínez	Universidad Autónoma Chapingo	México
Manuel Enrique Escanio Cortés	Universidad Nacional Autónoma de México	México
Víctor Hugo Sierra Nieto	Universidad Católica de Santiago de Guayaquil	Ecuador
Estela Maricela Villalón de la Isla	Centro Universitario del Norte de la Universidad de Guadalajara	México

María del Socorro Dena Medécigo	Universidad del Valle de México - Campus Hermosillo	México
Juan José Tacuri Quezada	Hospital IESS Quito Sur	Ecuador
Marvin Rodríguez Moreno	Universidad Nacional Autónoma de Honduras	Honduras
Diana Ereya Lugo Ondarza	Universidad Autónoma de Nuevo León	México
Ana Cecilia Ballesteros Payán	Universidad Autónoma de Sinaloa	México
Griselda Castro Contreras	Universidad de Morelos	México
Rafael Netzahualcóyotl Gallaga García	Universidad de Morelia	México
Juana Maygualidia Aguilar Gutiérrez	Instituto de Ciencias de la Salud de la Universidad Autónoma del Estado de Hidalgo	México
Sealtiel Armenta Arellano	Universidad Veracruzana	México
Irma Martínez González	Conalep Cuautitlán I. - Escuela de Enfermería General Mauro Soto	México
María Isabel Mendoza García	Universidad Veracruzana - Facultad de Medicina	México
Elva Jiménez Hernández	Universidad Nacional Autónoma de México	México
Martha Eunice Rodríguez Arellano	Hospital Regional Lic. Adolfo López Mateos - ISSSTE	México
Marco Antonio Román Villegas	Universidad Internacional de La Rioja	México
María de los Ángeles Ruiz Santiago	Investigador independiente	México
Oscar Humberto Oliva Chávez	Tecnológico de Estudios Superiores de Coacalco	México
Wilmer Ortega Chávez	Universidad Nacional Intercultural de la Amazonía	Perú

DEDICATORIA

A todos los participantes del masterclass, cuyo valioso aporte y contribución intelectual forman parte de la construcción de este libro, en el mismo sentido en que se reconoce la coautoría en la producción científica.

A ti, que abres las páginas de este libro y lo asumes como una herramienta para perfeccionar tu labor investigativa, con la convicción de que cada hallazgo y cada evidencia contribuirán a mejorar la sociedad.

AGRADECIMIENTOS ESPECIALES

Se expresa un reconocimiento a Javier Alonso Trujillo, Yurela Kosett Yunkor Romero y José Mario Ochoa Pachas, quienes en sus investigaciones han hecho referencia a las primeras formulaciones de los niveles de investigación.

Sus contribuciones han permitido que esta propuesta circule en la literatura antes de la presente sistematización, constituyendo un valioso antecedente para su consolidación en esta obra.

Véase, por ejemplo:

Alonso Trujillo J, Alonso Ricardez A. Aprendizaje estadístico basado en niveles de investigación. *Revista Educación*. 2022;46(1):1–16.

Yunkor Romero YK, Ochoa Pachas JM. Validación interna de instrumentos de investigación en las ciencias sociales. *Acta Jurídica Peruana*. 2020;3(2):112–35.

Ochoa Pachas JM, Romero Yunkor YK. El estudio descriptivo en la investigación científica. *Acta Jurídica Peruana*. 2019;2(2).

CONTENIDO

Parte I – Fundamentos de clasificación de los estudios	1
1. Introducción	3
2. Clasificación de los estudios	4
3. Antecedentes de los niveles de investigación	6
4. Los niveles de investigación según su propósito	9
5. Análisis de datos	12
5.1. Análisis de datos no estadístico	13
5.2. Análisis estadístico univariado	14
5.3. Análisis estadístico bivariado	15
5.4. Análisis estadístico multivariado	16
5.5. Modelamiento estadístico predictivo	17
5.6. Metodologías de mejora continua	18
Parte II – Niveles de investigación basado en el análisis de datos	21
1. Nivel exploratorio	23
1.1. La fenomenología	24
1.2. La hermenéutica	25
1.3. La heurística	26
1.4. El constructivismo	27
1.5. El estudio histórico	29
1.6. El estudio documental	30
1.7. El estudio del discurso	31
1.8. El estudio etnográfico	32
1.9. El estudio de caso	33
1.10. La teoría fundamentada	34
1.11. Objetivos del nivel exploratorio	36
1.12. Objetivo identificar (reconocer)	37
1.13. Objetivo interpretar (comprender)	38
1.14. Objetivo determinar (precisar)	39
1.15. Objetivo definir (operacionalizar)	40
2. Nivel descriptivo	41
2.1. La estadística en la investigación científica	42
2.2. Estudios no inferenciales	43
2.3. Estudios inferenciales	44
2.4. Objetivo describir o caracterizar	46
2.5. Objetivo estimar o calcular un parámetro	52
2.6. Objetivo contrastar o comparar con un parámetro	56
2.7. El ritual de la significancia estadística	57
3. Nivel relacional	61
3.1. Relación entre variables	62
3.2. La validez externa	63
3.3. Variables fijas y aleatorias	64
3.4. Objetivo comparar grupos independientes	65
3.5. Objetivo comparar medidas repetidas	68

3.6. Objetivo asociar y correlacionar	71
3.7. Objetivo medir la fuerza de asociación o correlación	74
3.8. El nivel relacional no es correlacional	78
4. Nivel explicativo	79
4.1. Criterios de causalidad	80
4.2. Análisis multivariado y causalidad	81
4.3. Variable independiente y dependiente	82
4.4. Objetivo evidenciar (influencia)	83
4.5. Procedimientos estadísticos multivariados	84
4.6. Análisis de la varianza (ANOVA)	85
4.7. Análisis estratificado de mantel-Haenszel	86
4.8. Regresión logística binaria	87
4.9. Objetivo demostrar (experimental)	88
4.10. Objetivo probar (consistencia)	93
5. Nivel predictivo	95
5.1. Relación entre variables	96
5.2. Modelos de ecuaciones estructurales (SEM)	97
5.3. Minería de datos (Data Mining)	98
5.4. Análisis de supervivencia	99
5.5. Series temporales	100
5.6. Objetivo predecir (SEM, DM)	102
5.7. Objetivo pronosticar (análisis de supervivencia)	103
5.8. Objetivo proyectar (series de tiempo)	104
5.9. Inferencia causal vs. modelamiento predictivo	106
6. Nivel aplicativo	107
6.1. Control estadístico de procesos (CEP)	108
6.2. Metodología Seis Sigma (DMAIC)	109
6.3. Metodología CRISP-DM	110
6.4. Objetivo evaluar (incluye monitorear)	113
6.5. Objetivo calibrar (incluye optimizar)	114
6.6. Objetivo controlar (incluye consolidar)	115
6.7. Objetivo aceptar (incluye validar)	116
6.8. La investigación tecnológica es parte del nivel aplicativo	118
Parte III – Derivados de los niveles de investigación	119
1. Derivados de los niveles de investigación	121
2. Marco conceptual general de las variables analíticas	122
3. Variables analíticas en el nivel exploratorio	124
4. Variables analíticas en el nivel descriptivo	125
5. Variables analíticas en el nivel relacional	126
6. Variables analíticas en el nivel explicativo	127
7. Variables analíticas en el nivel predictivo	128
8. Variables analíticas en el nivel aplicativo	129

PRESENTACIÓN

Hace algunos años tuve la necesidad de revisar libros de estadística de diversos autores, que en vez de despejar dudas, las complicaba, empleaba muchas horas de lectura y toma de apuntes, sin embargo no lograba consolidar un conocimiento sólido, llegaba al desánimo y la frustración; sin embargo una noche encontré un libro de Metodología de la Investigación que me atrapó por su forma de presentar a la investigación como una herramienta de manejo menos complicada, con bases metodológicas que comprendí perfectamente, con ejercicios vivenciales y la aplicación estadística que me resolvió la vida profesional. A partir de ahí, desde hace 12 años sigo preparándome y cada día me sorprendo de lo que se puede lograr en estadística e investigación científica.

En cada curso he logrado descubrir los objetivos planteados y resuelto todas las dudas que se presentan desde el inicio del estudio con la búsqueda de la línea de investigación, el enunciado del estudio, los objetivos, las variables y su operacionalización, las hipótesis y su aleatorización, así como el manejo estadístico hasta la interpretación de los resultados. El camino ha sido largo, de mucho aprendizaje, sobre todo de vastas experiencias hasta llegar a la consolidación del presente.

Niveles de Investigación es un libro donde el lector encontrará una clasificación emergente fundada en el análisis de datos para la investigación científica, producto de años de estudio realizados por el Dr. José Supo y compartidos con colegas de varios países, quienes han aportado al conocimiento científico, así como a su aplicación en los distintos ámbitos de la ciencia, tanto a las ciencias naturales como a las ciencias sociales.

El camino recorrido para llegar hasta esta propuesta metodológica plantea una clasificación de estudios que han sido aplicados desde una perspectiva operativa estableciendo criterios específicos al determinar cuál o cuáles niveles de investigación serían los más adecuados para su manejo estadístico. Las consideraciones que orientan la elección del enfoque metodológico son según el tipo de datos y las preguntas de investigación.

El investigador necesita poseer conocimientos sólidos en estadística para el manejo oportuno de las herramientas pertinentes para cada momento del estudio, además será capaz de planificar, operar, intervenir y evaluar cada una de las intervenciones que se realicen. Cada uno de los niveles de investigación tiene un objetivo fundamental que lo caracteriza según su profundidad o alcance, su finalidad, su enfoque y el lugar de obtención de datos, estos criterios fueron aportaciones de diversos autores.

Se presentan seis niveles de investigación: exploratorio, descriptivo, relacional, explicativo, predictivo y aplicativo.

La investigación exploratoria busca obtener información preliminar sobre un tema poco conocido, con el fin de identificar, clarificar o formular preguntas o hipótesis para investigaciones posteriores más profundas, pretende generar ideas iniciales y explorar las características generales del fenómeno.

La investigación descriptiva está orientada a observar, analizar y detallar las características particulares de un fenómeno, sujeto o situación tal como se presenta en la realidad, el investigador se limita a medir y exponer sistemáticamente características, propiedades o fenómenos observables en una población específica, con el fin de documentarlos y entender mejor su naturaleza y distribución, permiten identificar y precisar las características fundamentales de un grupo, acontecimiento o fenómeno, proporcionando información útil para otros tipos de investigaciones.

La investigación relacional se enfoca en analizar o buscar las relaciones entre dos o más variables, identificando cómo estas variables están conectadas entre sí, sin necesariamente implicar causalidad directa, busca comprender las asociaciones entre las variables implicadas. Existe una variable de supervisión y una o más variables asociativas, cuya relación respecto a la variable de supervisión es examinada individualmente o en conjunto para identificar patrones de relación.

La investigación explicativa busca esclarecer las causas que generan un fenómeno o situación particular, para encontrar relaciones causales o determinantes específicos

entre variables. Su objetivo principal es identificar y analizar las causas que originan un fenómeno determinado y cómo estas causas están relacionadas con sus efectos o consecuencias. Establece hipótesis que son sometidas a comprobación mediante la observación controlada y la aplicación de técnicas estadísticas para validar las relaciones causales propuestas. Exigen rigurosidad metodológica y un control riguroso de variables que permita aislar claramente la relación causal.

La investigación predictiva tiene como finalidad anticipar o pronosticar resultados futuros basándose en la relación entre variables observadas y medidas previamente. Se apoya en modelos estadísticos y matemáticos, permitiendo establecer patrones, tendencias o probabilidades sobre la ocurrencia de fenómenos específicos. Su validez y confiabilidad dependen de la precisión en la selección y operacionalización de variables, así como del uso adecuado de técnicas estadísticas y métodos de análisis cuantitativo. Su relevancia radica en la capacidad de generar conocimiento aplicable a contextos reales, facilitando la toma de decisiones basada en evidencias contribuyendo a la planificación y prevención en distintos ámbitos como la salud pública, la economía, la educación, entre otros.

La investigación aplicada está orientada a resolver problemas prácticos o a transformar positivamente la realidad. Su objetivo principal es producir cambios concretos mediante la implementación de soluciones, intervenciones o innovaciones tecnológicas que respondan a necesidades específicas identificadas en la sociedad. Se divide en dos subtipos: la investigación sustantiva, que busca crear productos tangibles como prototipos, y la investigación operativa, relacionada con la optimización de sistemas y redes. Se caracterizan por generar resultados aplicables directamente a la mejora de procesos, servicios o productos, y no solo teorías o explicaciones científicas abstractas.

En cada nivel de investigación se explica a detalle la fundamentación teórica, las características y requisitos a considerar para determinar su aplicabilidad, la posibilidad de intervención o no intervención, el planteamiento de la hipótesis, el análisis estadístico, el ritual de la significancia y la interpretación de los resultados, transitar por cada nivel permite determinar si se continua al siguiente.

El autor del libro y los coautores invitan al lector a sumergirse por los amplios caminos de la investigación y el manejo estadístico de los estudios que vayan surgiendo en su quehacer cotidiano.

Dra. María Martina Kantún Can
Senadora de la República por el Estado de Campeche, México
Doctora en Educación Humanista, Instituto Humanista de Sinaloa
Ex Subsecretaria de Educación Básica y Normal, Estado de Campeche, México

PRÓLOGO

Escribir estas líneas para presentar la obra *Niveles de Investigación: Una clasificación emergente fundamentada en el análisis de datos para la investigación científica* representa para mí un honor y un compromiso profundamente personal. El Dr. José Supo, además de ser un académico brillante y un referente en el campo de la bioestadística aplicada a la investigación científica, ha sido para mí maestro, guía y ejemplo de rigor intelectual. Sus enseñanzas han marcado un rumbo decisivo en mi formación metodológica y en mi quehacer académico. Sus lecciones no se limitaron a la transmisión de conocimientos técnicos: fueron, sobre todo, una invitación a pensar críticamente, a formular preguntas con precisión y a buscar respuestas con disciplina y ética, inspirándome a acompañar con solidez y claridad los procesos de investigación de las nuevas generaciones de médicos residentes.

La presente obra constituye mucho más que un manual de clasificación metodológica: es el resultado de años de experiencia, reflexión crítica e innovación en la manera de entender el análisis de datos como columna vertebral de la investigación científica.

El libro se estructura por partes. La primera presenta los fundamentos de la clasificación, abordando la evolución histórica del concepto, la diversidad de criterios utilizados por distintos autores y, sobre todo, la centralidad del análisis de datos en la determinación del nivel de investigación. En este recorrido, el lector se adentra en un espectro que va desde métodos no estadísticos, como la fenomenología, la hermenéutica o el constructivismo, hasta técnicas estadísticas univariadas, bivariadas y multivariadas, el modelamiento estadístico y las metodologías de mejora continua.

La segunda parte desarrolla, uno por uno, los seis niveles de investigación, precisando su propósito, alcance y las herramientas analíticas que los caracterizan. El nivel exploratorio abre el camino para comprender fenómenos poco conocidos; el descriptivo permite caracterizar con detalle variables y poblaciones; el relacional explora asociaciones significativas; el explicativo profundiza en las relaciones causales; el predictivo anticipa resultados futuros mediante modelos robustos; y el aplicativo orienta

la ciencia hacia la solución de problemas concretos y la optimización de procesos.

Este enfoque progresivo, que articula el método científico con el análisis de datos, ofrece al investigador un mapa claro para transitar desde la exploración inicial hasta la aplicación práctica de los hallazgos. Más aún, el libro fomenta una visión integradora: la investigación no es un proceso lineal y aislado, sino un ciclo que se retroalimenta, en el que cada nivel se nutre de los anteriores y prepara el terreno para el siguiente.

Como médico Intensivista Pediatra y académica, reconozco en esta obra no solo un compendio teórico de gran valor, sino también una herramienta formativa que contribuirá a mejorar la calidad, la coherencia y la aplicabilidad de la investigación científica en nuestro contexto latinoamericano y global. El rigor conceptual del Dr. José Supo Condori y la experiencia interdisciplinaria de sus coautores se reflejan en cada página, haciendo de este texto una referencia obligada para estudiantes, docentes e investigadores que buscan llevar sus proyectos a un estándar metodológico más alto.

Por todo ello, invito al lector a adentrarse en estas páginas con la misma apertura y entusiasmo con que yo las he recorrido, seguro de que encontrará aquí una fuente de claridad metodológica y de inspiración académica.

Dra. Ma. Elizabeth Rábago Sánchez

Médico Intensivista Pediatra

Doctora en Ciencias Médicas

Profesora Investigadora, Universidad Juárez del Estado de Durango, México

**PARTE I - FUNDAMENTOS DE
CLASIFICACIÓN DE LOS
ESTUDIOS**

1. INTRODUCCIÓN

La evolución reciente de la estadística aplicada y de la ciencia de datos exige una redefinición metodológica que facilite la integración de estas herramientas en la investigación científica. La estadística aplicada ha sido un pilar para sintetizar, interpretar y validar datos, mientras que la ciencia de datos, al combinar estadística avanzada, programación y aprendizaje automático, abre nuevas posibilidades para modelar fenómenos complejos, identificar patrones y generar predicciones.

La propuesta metodológica aquí presentada plantea una clasificación de estudios que, desde una perspectiva operativa, establece criterios claros para determinar y aplicar el nivel de investigación más adecuado. Esto implica definir fronteras instrumentales que orienten la elección de enfoques metodológicos según el tipo de datos y las preguntas de investigación.

Beneficios de la clasificación por análisis de datos:

- Facilita la selección de herramientas analíticas congruentes con el nivel investigativo.
- Mejora la reproducibilidad de los estudios al definir criterios operativos claros.
- Permite la integración de técnicas modernas de minería de datos, modelamiento estadístico y metodologías de mejora continua.
- Optimiza la planificación y evaluación de intervenciones basadas en evidencia.

Limitaciones actuales que esta propuesta busca superar:

- Clasificaciones previas centradas en criterios teóricos y cualitativos, con baja precisión operativa.
- Escasa compatibilidad con procedimientos estadísticos avanzados y análisis de datos masivos.
- Falta de criterios para abordar fenómenos complejos que requieren análisis multivariado o predictivo.

2. CLASIFICACIÓN DE LOS ESTUDIOS

No existe un consenso único sobre cómo “debe” clasificarse un estudio y distintos autores priorizan criterios diferentes, como instrumentos para planificar, ejecutar y comunicar estudios con rigor, de modo que todas son metodológicas.

Según su profundidad o alcance, explorar cuando el fenómeno es difuso; describir cuando se necesita caracterizar; correlacionar cuando interesa las asociaciones; y explicar cuando se buscan condiciones causales. Esta lógica de clasificación se consolidó en la tradición metodológica de Selltiz, se difundió en los manuales de Babbie y fue sistematizada por Dahnke al incorporar explícitamente el nivel exploratorio (1).

Según su finalidad, diferencia entre investigación básica o pura, centrada en ampliar conocimiento teórico y afinar conceptos; e investigación aplicada, orientada a resolver problemas concretos con ese conocimiento. La investigación básica, además de validez interna busca aporte conceptual; La investigación aplicada busca efecto demostrable, costo-efectividad Tamayo (2) y Arias (3).

Según su enfoque, el enfoque cuantitativo privilegia la medición, la estimación y la inferencia estadística; el cualitativo busca comprensión e interpretación contextual; y el mixto integra ambos cuando la complejidad del problema lo exige (dice la teoría). En la práctica, elegir enfoque es elegir cómo se construirá la evidencia: tamaños muestrales y métricas de precisión en lo cuantitativo; saturación teórica, triangulación y trazabilidad interpretativa en lo cualitativo Denzin & Lincoln (4), Creswell & Plano Clark (5).

Según el lugar de obtención de datos, el documental trabaja con fuentes registradas y exige protocolos de búsqueda y evaluación de calidad; lo de campo implica capturar datos primarios con estándares de medición, muestreo y ética, cuando es viable. Lo experimental, cuando procede, permite estimar efectos causales bajo supuestos más estrictos Arias (3).

Cuadro 1: Clasificación de los estudios según criterios metodológicos

Criterio	Definición	Categorías principales	3 autores clave
Por profundidad o alcance	Clasifica el estudio según el nivel de detalle y análisis del fenómeno observado.	<ul style="list-style-type: none"> - Exploratorio: indaga fenómenos poco conocidos. - Descriptivo: caracteriza fenómenos. - Correlacional: examina asociaciones. - Explicativo: busca causas. 	Selltiz, C. (1959) Kerlinger F. (1964) Dahnke, G. (1986)
Por finalidad	Diferencia el propósito teórico o práctico de la investigación.	<ul style="list-style-type: none"> - Básica o pura: expande el conocimiento teórico. - Aplicada: resuelve problemas concretos con base en conocimiento existente. 	Bernal J. D. (1939) Bush V. (1945) Bunge M. (1960)
Por enfoque	Define el marco epistemológico y las técnicas predominantes para generar evidencia.	<ul style="list-style-type: none"> - Cuantitativo: medición y análisis estadístico. - Cualitativo: comprensión contextual e interpretativa. 	Durkheim E. (1895) Weber M. (1904) Glaser B. (1967)
Por lugar de obtención de datos	Clasifica según la fuente y contexto de recolección de información.	<ul style="list-style-type: none"> - Documental: usa fuentes registradas. - De campo: recolecta datos primarios. 	Malinowski B. (1922) Allport G. (1942) Good C. (1954)

3. ANTECEDENTES DE LOS NIVELES DE INVESTIGACIÓN

Ander-Egg (6), plantea que la investigación científica se puede clasificar en esquemas o **niveles** y señala que es una intención muy generalizada, tanto en el ámbito de las ciencias sociales como en el de las ciencias físicas, aunque el autor distingue únicamente tres niveles de investigación (descripción, clasificación y explicación), agrega que cada uno de estos niveles indica diversos grados de profundidad y, en consecuencia, diferentes exigencias y dificultades metodológicas. Del contenido del autor podemos deducir que la “clasificación” correspondería al nivel descriptivo, aunque únicamente para el objetivo caracterizar.

Para Carpio Valencia (7), existen también **niveles** (exploratoria, descriptiva, correlacional, explicativa y predictiva) que pueden conseguir las investigaciones y para desarrollarlos se requiere un tipo de análisis, con lo que el autor está confirmando que el mejor argumento para definir operativamente los niveles de investigación es su tipo de análisis. Adicionalmente el autor no precisa cuáles son esos análisis que distinguen un nivel de otro, con lo cual nuevamente nos encontramos ante la ausencia de una división operativa, agregando además que estos dos primeros autores no coinciden en la identificación de los niveles.

Por otro lado, Alonso Trujillo (8), resalta que el concepto de **niveles** de investigación, basado en el análisis de datos es el criterio más adecuado, incluso desarrolló un estudio de aprendizaje estadístico basado en niveles de investigación, cuyo propósito fue mejorar el aprendizaje de la estadística en estudiantes de la licenciatura en enfermería en una institución educativa de nivel superior en México, para su estudio consideró los cuatro primeros niveles: exploratorio, descriptivo, relacional y explicativo, siendo que el primero se desarrolla sin análisis estadístico, su clasificación correspondió al análisis estadístico univariado, bivariado y multivariado. Su propuesta constituye una valiosa contribución a la integración de la metodología de investigación con el aprendizaje estadístico, y abre el camino para futuras aplicaciones que incorporen niveles metodológicos de mayor complejidad, como el predictivo y el aplicativo.

Supo Condori (9), Sostiene que toda línea de investigación transita dos fases: una fase cualitativa, el nivel exploratorio, que aclara el fenómeno sin uso de la estadística, y una fase cuantitativa que comprende los **niveles** descriptivo, relacional, explicativo, predictivo y aplicativo, donde la bioestadística se convierte en herramienta central. Aquí se plantea: el nivel descriptivo, con análisis estadístico univariado, el nivel relacional que trabaja con estadística bivariada sin atribuir causalidad. El nivel explicativo incorpora análisis multivariado y, cuando es pertinente, diseños experimentales, el nivel predictivo que recurre a modelos de predicción para anticipar resultados, y finalmente, el nivel aplicativo con análisis estadístico orientado al control de la calidad y la mejora continua.

Avellaneda Callirgos (10), por su lado presenta los seis **niveles** de investigación: exploratorio, descriptivo, relacional, explicativo, predictivo y aplicativo, como una metodología ágil, utilizando palabras clave para su diferenciación, sin presentar mayor sustento o argumento, sin embargo al ser un autor contemporáneo vemos que el concepto en sí ya está difundido, mas no su sustento, ni los argumentos que llevan a presentar esta clasificación, más allá de su alcance, enfoque o profundidad, como mencionan muchos autores, dado que el aprendizaje automático y la ciencia de datos requiere de una clasificación operativa.

Yunkor Romero (11) identifica los seis **niveles**: exploratorio, descriptivo, relacional, explicativo, predictivo y aplicativo. Con el fin de ejecutar validación interna de instrumentos de investigación en las ciencias sociales, para lo cual se apoya en esta clasificación, lo cual resulta bastante conveniente, sin embargo, en el cuerpo del documento, no se pretende dar sustento de las estrategias que pueden ayudar a validar instrumentos basado en el análisis de datos, por cuanto aún no se cuenta con un cuerpo teórico del análisis de datos para cada nivel investigativo.

El consenso de todos estos autores radica en que la clasificación de los estudios en **niveles** de investigación facilita al investigador seleccionar estrategias metodológicas apropiadas según la finalidad específica de su estudio, proporcionando coherencia lógica y operativa a todo proceso investigativo.

Cuadro 2. Aparición cronológica de los niveles de investigación

Año	Autor	Niveles expuestos	Fundamento
1959	J. S. Seltiz	Exploratorios, Descriptivos y Explicativos	Metodológico
1964	Kerlinger, Fred	Exploratorios, Descriptivos y Experimentales	Metodológico
1986	Dahnke, Gordon	Exploratorios, descriptivos, correlacionales y explicativos	Metodológico
2011	Ander-Egg	Descriptivos, Clasificatorios y Explicativos	Metodológico
2013	Supo Condori	Exploratorio, Descriptivo, Relacional, Explicativo, Predictivo y Aplicativo	Análisis de datos
2018	Carpio Valencia	Exploratorio, Descriptivo, Correlacional, Explicativo y Predictivo	Análisis de datos
2020	Yunkor Romero	Exploratorio, Descriptivo, Relacional, Explicativo, Predictivo y Aplicativo	Análisis de datos
2022	Avellaneda Callirgos	Exploratorio, Descriptivo, Relacional, Explicativo, Predictivo y Aplicativo	Análisis de datos
2022	Alonso Trujillo	Exploratorio, Descriptivo, Relacional y Explicativo	Análisis de datos

4. LOS NIVELES DE INVESTIGACIÓN SEGÚN SU PROPÓSITO

La investigación exploratoria se orienta a reunir datos preliminares sobre un tema poco conocido, con el fin de generar preguntas o hipótesis que guíen indagaciones más avanzadas. Este tipo de estudio es útil especialmente cuando la información previa o disponible no es suficiente, permitiendo al investigador indagar sobre un área temática específica y orientando futuras investigaciones hacia aspectos más definidos y específicos. La investigación exploratoria no pretende ofrecer resultados concluyentes, sino más bien generar ideas iniciales y explorar las características generales del fenómeno que serán luego abordadas con mayor detalle en etapas posteriores del proceso investigativo (12).

La investigación descriptiva tiene por objetivo observar y documentar con precisión los rasgos particulares de un fenómeno, sujeto o situación, tal como se presentan en la práctica, sin alterar las variables ni pretender identificar relaciones de causa y efecto (6,10). En un estudio descriptivo, el investigador se dedica únicamente a describir de forma ordenada los rasgos y fenómenos observables en una población, con el objetivo de documentarlos y entender su distribución (2, 3). Estos estudios permiten identificar y precisar las características fundamentales de un grupo, acontecimiento o fenómeno, proporcionando información útil para otros tipos de investigaciones, como las relacionales, explicativas o predictivas (10,13).

La investigación relacional se enfoca en analizar y buscar las relaciones entre variables, identificando cómo estas variables están conectadas entre sí, sin necesariamente implicar causalidad directa. Este nivel investigativo es más avanzado que el descriptivo, dado que no solamente se limita a la observación y registro de las características o fenómenos, sino que busca comprender las asociaciones entre las variables implicadas (7). En términos generales, en el nivel relacional, hay una variable de supervisión (considerada la variable central del análisis) y una o más variables asociativas (vinculantes), cuya relación con la variable de supervisión se evalúa individual o conjuntamente para reconocer patrones.

La investigación explicativa busca determinar las causas que provocan un fenómeno o resultado, yendo más allá de la simple descripción o asociación para encontrar relaciones causales o determinantes específicos entre variables (10). Este tipo de investigación tiene como objetivo principal identificar, describir y analizar las causas que originan un fenómeno determinado y cómo estas causas están relacionadas con sus efectos o consecuencias (7). Para ello, establece hipótesis que son sometidas a comprobación mediante la observación controlada y la aplicación de técnicas estadísticas para validar las relaciones causales propuestas (8). Además, estos estudios exigen rigurosidad metodológica y un control riguroso de variables que permita aislar claramente la relación causal (14).

La investigación predictiva tiene como finalidad anticipar o pronosticar resultados futuros basándose en la relación entre variables observadas y medidas previamente. Este tipo de estudio se apoya en modelos estadísticos y matemáticos, permitiendo establecer patrones, tendencias o probabilidades sobre la ocurrencia de fenómenos específicos (13,14). La validez y la confiabilidad se sustentan en la correcta selección y adecuada operacionalización de las variables, así como en la aplicación rigurosa de técnicas estadísticas y métodos de análisis cuantitativo (13,14). Su valor principal está en producir conocimiento aplicable que oriente decisiones con sustento empírico y favorezca la planificación y prevención en sectores como salud, economía y educación.

La investigación aplicada se orienta a generar soluciones prácticas a problemas concretos y a impulsar transformaciones positivas en la realidad social, económica o institucional. Su propósito es producir cambios observables y verificables mediante propuestas, programas o innovaciones que respondan a necesidades específicas de la comunidad o del entorno profesional. A diferencia de la investigación básica, que se concentra en la construcción teórica y explicativa abstracta, la aplicada se distingue por ofrecer resultados con utilidad inmediata y relevancia contextual. Estos estudios contribuyen al perfeccionamiento de procesos, a la optimización de servicios y a la mejora de productos (3,15).

Cuadro 3. Tipos de investigación según su propósito

Tipo de investigación	Propósito principal	3 autores primarios
Exploratoria	Obtener información preliminar sobre fenómenos poco conocidos para identificar variables, categorías o hipótesis iniciales; no busca resultados concluyentes.	Malinowski B. (1922) Glaser B. (1967) Tukey J. (1977)
Descriptiva	Caracterizar sistemáticamente fenómenos, sujetos o contextos, midiendo y documentando sus atributos sin establecer relaciones causales.	Mead M. (1928) Kinsey A. (1948) Geertz C. (1973)
Relacional	Examinar asociaciones entre variables (una de supervisión y otras asociativas), cuantificando dirección e intensidad del vínculo sin atribuir causalidad.	Spearman C. (1904) Moreno J. (1934) Cronbach L. (1957)
Explicativa	Establecer y comprobar relaciones causales entre variables, sometiendo hipótesis a verificación estadística u observacional.	Fisher R. (1935) Campbell D. (1963) Hill A. (1965)
Predictiva	Anticipar resultados futuros mediante modelos estadísticos o de aprendizaje automático, con énfasis en su validación interna y externa (calibración y discriminación).	Meehl P. (1954) Forrester J. (1961) Box & Jenkins. (1970)
Aplicativa	Implementar y evaluar intervenciones o innovaciones que resuelvan problemas prácticos, consolidando mejoras sostenibles.	Taylor F. (1911) Lewin K. (1946) Stokes D. (1997)

5. ANÁLISIS DE DATOS

El análisis de datos es la fase que enlaza la recolección de información con la explicación científica. Consiste en inspeccionar, interpretar y extraer conclusiones a partir de los datos recopilados, con o sin el uso de la estadística. Una vez obtenidos los datos, el investigador debe organizarlos, depurarlos y aplicar técnicas apropiadas para determinar si los resultados respaldan la hipótesis planteada.

El análisis de datos. En la investigación empírica se recurre a la estadística descriptiva para explorar y resumir la información y a la estadística inferencial para contrastar hipótesis y generalizar los hallazgos. Estas herramientas proporcionan una alternativa objetiva al juicio personal y ayudan a evitar conclusiones erróneas. Además, el análisis de datos no se limita al uso de la estadística: combinar herramientas cualitativas y cuantitativas, por ejemplo, codificación manual de entrevistas o uso de software especializado, amplía la comprensión de los fenómenos estudiados.

Relación con el método científico. Dentro del método científico, el análisis de datos se sitúa tras la experiencia y antes de la conclusión. Una vez recogidos los datos se aplican técnicas de análisis estadísticas y no estadísticas para interpretar los resultados. El rigor con el que se organiza, limpia y analiza la información influye directamente en la validez de las conclusiones; por ello, las guías metodológicas subrayan que el procesamiento y análisis de datos son pasos esenciales del procedimiento científico, al mismo nivel que la recolección o el diseño del estudio.

Importancia en la investigación científica. El análisis de datos transforma las observaciones en evidencias que sustentan o refutan hipótesis. Permite garantizar la validez y objetividad de los hallazgos al verificar la confiabilidad de los instrumentos y la pertinencia de las pruebas empleadas. Asimismo, orienta el diseño y la clasificación de la investigación, pues la elección de la estrategia metodológica depende del propósito del estudio y del análisis de datos que este requiere. En áreas aplicadas los resultados analíticos proporcionan bases sólidas para la toma de decisiones.

5.1. ANÁLISIS DE DATOS NO ESTADÍSTICO

Enfoque fenomenológico, como el que se observa en la etnografía centrada en una recopilación y análisis inductivo-contextual, destacando la perspectiva interna (émica) de los participantes. Para esto se requiere de una observación participante prolongada, complementada con entrevistas abiertas en el contexto natural, con una exigente reducción fenomenológica o epoché para minimizar prejuicios del investigador. Mantener una flexibilidad metodológica con análisis iterativos que permitan la emergencia inductiva de temas relevantes.

Enfoque hermenéutico, como el encontrado en el estudio del discurso, con un análisis interpretativo profundo de textos y contextos socioculturales, mediante un ciclo hermenéutico (de lo particular al contexto total y viceversa). Esta herramienta considera activamente la subjetividad del investigador como herramienta analítica, reconociendo conocimientos previos y sesgos. Utiliza métodos cualitativos flexibles que reflejen las múltiples realidades interpretativas construidas socialmente a través del lenguaje.

Enfoque heurístico, como lo que se detecta en el estudio de caso, el cual requiere de una exploración profunda para la generación de nuevos conocimientos. Utiliza casos específicos para generar hipótesis emergentes, adoptando un enfoque inductivo, describiendo detalladamente los casos para descubrir significados imprevistos. Sus conclusiones son analíticas o conceptuales, la documentación resalta cómo el conocimiento se va construyendo durante la investigación misma.

Enfoque constructivista, observado en la teoría fundamentada, combina observación y ensayo para llegar a conclusiones que, emergen desde los hechos de forma natural, su codificación inductiva es abierta. Promueve la reflexividad del investigador mediante memos analíticos para documentar la co-construcción del conocimiento. El producto es una teoría interpretativa y situada, donde el investigador reconoce su influencia en el análisis y construcción teórica final. Aunque todos estos estudios pueden incluir datos numéricos y categóricos, su análisis es netamente cualitativo.

5.2. ANÁLISIS ESTADÍSTICO UNIVARIADO

Analiza los datos estadísticamente de uno en uno, datos que provienen de la medición de las variables en el estudio en curso. Utiliza herramientas estadísticas univariadas diferenciadas para variables numéricas y categóricas. Las variables numéricas se describen con: medidas de tendencia central (media, mediana, moda), medidas de dispersión (rango, desviación estándar, varianza) medidas de posición (percentiles, cuartiles y deciles) y medidas de forma (curtosis y asimetría). Las variables categóricas se describen con: frecuencias (absolutas, relativas y acumuladas), razones (odds) y proporciones (tasas). Este nivel permite diferenciar dos momentos:

Momento no inferencial, netamente descriptivo, lo cual implica caracterizar a las unidades de estudio, sin la intención de llevar sus conclusiones hacia un conjunto mayor. En este nivel, la variable de estudio se denomina variable de interés y las demás se denominan variables de caracterización, la descripción estadística se aplica a todas las variables según su naturaleza.

Momento inferencial, aplica únicamente a la variable de estudio y del análisis realizado se plantea llevar las conclusiones extraídas de un conjunto denominado muestra, hacia un conjunto mayor denominado población esta inferencia se puede conseguir de dos formas: la estimación puntual y la prueba de hipótesis.

- En el caso de la estimación puntual: Cuando los datos son numéricos, la estimación puntual consiste en interpretar la media de la muestra, como media poblacional y calcular el intervalo de confianza. Cuando los datos son categóricos se interpretan con la frecuencia de la cualidad en estudio como prevalencia de la población.

- En el caso de la prueba de hipótesis, cuando la variable de estudio es numérica se desarrolla t de Student para una sola muestra, cuando la variable de estudio se desarrolla el χ^2 (ji-cuadrado) de bondad de ajuste, en ambos casos para contrastar el valor calculado en la muestra con parámetro establecido.

5.3. ANÁLISIS ESTADÍSTICO BIVARIADO

En este nivel se identifica una variable de supervisión como eje del análisis, y las demás se consideran variables asociativas. Explorar relaciones entre variables significa vincular cada variable asociativa (vinculante) con la variable de supervisión (central) para estudiar sus conexiones. Los procedimientos estadísticos dependen de la naturaleza de las variables ya sean categóricas o numéricas. Las variables categóricas pueden ser fijas o aleatorias, mientras que las variables numéricas siempre son aleatorias.

- Primer escenario, una variable fija y otra aleatoria, desarrolla el χ^2 (ji-cuadrado) de homogeneidad cuando la variable aleatoria es categórica, o la t de Student para muestras independientes cuando la variable aleatoria es numérica.
- Segundo escenario, ambas variables son aleatorias, desarrolla el χ^2 (ji-cuadrado) de independencia cuando las dos variables son categóricas y la correlación de Pearson cuando las dos variables son numéricas.
- Tercer escenario, las dos variables son fijas, siendo que solamente las variables categóricas pueden ser fijas, se entiende que ambas son categóricas, desarrolla la prueba exacta de Fisher si las variables categóricas son dicotómicas.

Si la intención va más a la demostración de la dependencia probabilística, se procede a realizar las medidas de asociación o las medidas de correlación según corresponda:

- Si existe asociación, se procede al cálculo de las medidas de asociación, que pueden ser de corte matemático o epidemiológico como el Riesgo Relativo (RR) o el Odds Ratio (OR), también aplica el Kappa de Cohen como medida de concordancia.
- Si existe correlación, se procede al cálculo de las medidas de correlación como la r de Pearson, o procedimientos no paramétricos como la Rho de Spearman, al caso de no contar con distribución normal, finalmente se puede desarrollar correlaciones para variables categóricas ordinales con la tau-b de Kendall o tau-c de Kendall.

5.4. ANÁLISIS ESTADÍSTICO MULTIVARIADO

Analizar a las variables de forma multivariada tiene como intención principal descubrir relaciones reales, verdaderas o causales, dicho de otro modo, descartar relaciones causales, aleatorias o espurias. Siendo que la dependencia probabilística es el principal criterio de causalidad, el análisis multivariado es parte del control de los estudios en los estudios de causa y efecto, su finalidad es controlar los posibles efectos de variables intervinientes, para luego complementar con otros criterios de causalidad. Aquí se pueden diferenciar dos momentos, uno sin intervención llamado observacional y otro con intervención llamado experimental.

Momento observacional (sin intervención), incluye a los estudios epidemiológicos, pudiendo ser prospectivos o retrospectivos, con una variable dependiente categórica o numérica. El diseño de casos y controles es explicativo si cuenta con análisis multivariado, el diseño de cohortes incorpora un fundamento causal clave: la relación temporal, ya que se estructura como un estudio prospectivo y longitudinal.

Si bien ambos diseños son observacionales, es decir no hay una variable manipulada, a efectos de desarrollar un estudio explicativo es imperativo señalar cuál variable es la posible causa. Para que, en el análisis estadístico, la variable propuesta como causa, sean controlada estadísticamente mediante análisis multivariado.

Momento experimental, (con intervención) incluye la manipulación de una o más variables independientes, aquí el control estadístico es ineludible, el experimento en su versión pura es prospectivo, y la variable dependiente numérica a fin de maximizar la probabilidad de encontrar una relación causal cuando esta exista.

Los procedimientos estadísticos incluyen al análisis de varianza (ANOVA) y todos sus sucedáneos, conocidos también como el modelo lineal general (MLG), como el diseño de bloques completamente aleatorizado (DBCA) que incluye una variable manipulada y un factor fijo denominado control, el diseño cuadrado latino (DCL), etc.

5.5. MODELAMIENTO ESTADÍSTICO PREDICTIVO

Predecir con Regresión Múltiple: La regresión múltiple permite predecir eventos o condiciones de salud combinando múltiples predictores (variables exógenas) para estimar el resultado esperado (variable endógena). Se utilizan regresiones logísticas y lineales para desarrollar modelos predictivos que permitan anticipar riesgos individuales y poblacionales. Esta técnica se validará externamente para garantizar su precisión predictiva, especialmente útil por su simplicidad operacional, así como por su adaptabilidad a datasets limitados.

Predecir con Minería de Datos: La precisión predictiva se consigue con técnicas avanzadas de minería de datos (*machine learning*) como árboles de decisión, bosques aleatorios y redes neuronales artificiales, aprovechando grandes volúmenes de datos clínicos y poblacionales. Estos modelos permiten predecir eventos críticos como brotes epidémicos, riesgo individual de enfermedades crónicas y probabilidades de reingresos hospitalarios. La ventaja operativa es la capacidad de manejo de relaciones no lineales.

Pronosticar con Análisis de Supervivencia: Utilizando técnicas como Kaplan-Meier y modelos de Cox, el análisis de supervivencia aportará pronósticos sobre el tiempo esperado hasta eventos de interés (muerte, recaídas, complicaciones). Operativamente, este método permitirá evaluar pronósticos individuales y grupales, identificando factores de riesgo y estableciendo curvas de supervivencia ajustadas por características clínicas y demográficas. Útil para proyectar la evolución clínica y planificar recursos de salud a mediano y largo plazo.

Proyectar con Series de Tiempo: La implementación de análisis de series temporales, especialmente con modelos SARIMA, permitirá proyectar eventos de salud como incidencia de enfermedades infecciosas y demanda de servicios médicos. Detectando tendencias y estacionalidades en los datos históricos permite anticipar necesidades futuras. Facilita la planificación en salud pública, permitiendo ajustes oportunos en recursos hospitalarios, optimizando la respuesta sanitaria y la asignación de recursos.

5.6. METODOLOGÍAS DE MEJORA CONTINUA

El Control Estadístico de Procesos (CEP) permite monitorear y optimizar procesos mediante análisis de variaciones en series temporales. Adaptado al contexto sanitario, el CEP ayuda a diferenciar la variabilidad natural de los procesos biológicos y organizacionales de aquellas debidas a causas específicas o asignables. Se emplea en la calibración de instrumentos e indicadores, optimización de procesos clínicos y validación estadística de métodos diagnósticos y flujos de atención sanitaria. A través de gráficos de control, permite un seguimiento continuo de métricas clave, facilitando la identificación temprana de desviaciones significativas que requieren intervención inmediata.

La metodología Seis Sigma, particularmente el enfoque DMAIC (Definir, Medir, Analizar, Mejorar y Controlar), proporciona una estructura para abordar problemas específicos mediante una secuencia lógica y estadísticamente rigurosa. En la fase de Definir se establece claramente el problema y objetivos específicos de mejora. En Medir, se recopilan datos precisos para caracterizar el desempeño inicial del proceso. La fase Analizar identifica las causas raíz de la variabilidad observada, mientras que Mejorar implica implementar soluciones específicas validadas estadísticamente a través de pruebas piloto. Finalmente, Controlar garantiza el mantenimiento sostenido de las mejoras mediante monitoreo continuo.

La metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*) ofrece un marco estructurado para proyectos de ciencia de datos aplicado a salud pública, dividido en seis fases iterativas: Comprensión del contexto, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue. Este enfoque permite calibrar modelos predictivos y analíticos ajustados a contextos epidemiológicos específicos, optimizar su precisión mediante ingeniería de características y ajustes metodológicos, validar resultados robustamente con datos históricos y pruebas prospectivas, y monitorizar continuamente su desempeño y utilidad en la práctica clínica o epidemiológica.

Cuadro 4. Tipo de análisis de datos

Propósito principal	Métodos y técnicas principales	Tipo de análisis de datos	Nivel emergente
Identificar y definir fenómenos o conceptos emergentes	Fenomenología, hermenéutica, estudio de caso, teoría fundamentada	Análisis de datos no estadístico	Exploratorio
Describir variables individualmente y caracterizar unidades de estudio	Media, mediana, moda, desviación estándar, frecuencias, t de Student, χ^2 bondad de ajuste.	Análisis estadístico univariado	Descriptivo
Explorar relaciones entre dos variables (supervisión y asociativas)	χ^2 de homogeneidad/independencia, t de Student, Pearson, OR, RR	Análisis estadístico bivariado	Relacional
Explicar relaciones de causalidad controlando variables intervinientes	ANOVA, MLG, estudios de caso control, estudios de cohortes y experimentales	Análisis estadístico multivariado	Explicativo
Predecir el comportamiento de una variable a partir de otras	Regresión múltiple, árboles de decisión, redes neuronales, Kaplan-Meier, SARIMA	Modelamiento estadístico predictivo	Predictivo
Optimizar procesos mediante intervención y monitoreo continuo	CEP, Seis Sigma (DMAIC), CRISP-DM, gráficos de control, índices Cp y Cpk	Metodologías de mejora continua	Aplicativo

Recorrido de una línea de investigación a través de los niveles

Una **línea de investigación** es un eje temático que acumula conocimiento sobre un mismo objeto de estudio. Su recorrido metodológico transita por los niveles de investigación: inicia explorando lo desconocido, describe sus características, constata relaciones, explica causas, predice escenarios y, finalmente, aplica intervenciones para transformar la realidad.

Cuadro 5. Ejemplo aplicado: SARS-CoV-2 y COVID-19 en trabajadores de salud.

Nivel	Pregunta que responde	Objeto de análisis	Ejemplo
Exploratorio	¿Qué se necesita identificar o comprender?	Fenómenos o significados	Percepciones de trabajadores de salud sobre el diagnóstico de infección por SARS-CoV-2
Descriptivo	¿Qué características o frecuencias presenta el objeto de estudio?	Variables individuales	Incidencia de COVID-19 en trabajadores de salud durante la primera ola
Relacional	¿Qué vínculos existen entre las variables?	Relaciones entre variables	Asociación entre obesidad y COVID-19 en trabajadores de salud
Explicativo	¿Qué factores influyen en su ocurrencia o variación?	Relaciones causales	Determinantes de la progresión a enfermedad en trabajadores de salud con infección por SARS-CoV-2
Predictivo	¿Qué resultados pueden anticiparse a partir de los datos disponibles?	Modelos predictivos	Modelo predictivo de ingreso a cuidados intensivos en trabajadores de salud con COVID-19
Aplicativo	¿Qué acciones permiten intervenir, controlar o mejorar la situación?	Resultados de intervención	Efectividad de la vacunación contra SARS-CoV-2 en la reducción de hospitalizaciones y mortalidad por COVID-19 en trabajadores de salud

**PARTE II - NIVELES DE
INVESTIGACIÓN BASADO EN
EL ANÁLISIS DE DATOS**

NIVEL

EXPLORATORIO

ANÁLISIS DE DATOS NO ESTADÍSTICO

Los estudios de nivel exploratorio indagan fenómenos poco estudiados y a descubrir dimensiones inesperadas de la realidad, reconociendo la posibilidad de hallazgos fortuitos o serendípicos. Su propósito es generar una comprensión inicial del objeto de estudio mediante la identificación de significados, percepciones y contextos.

En este nivel la atención recae en la entidad de interés, con el fin de reconocer sus características fundamentales y configurarla como variable de estudio. Esta configuración inicial permite darle forma analítica al fenómeno, dotándolo de una identidad investigativa que servirá de base para posteriores niveles de análisis.

A diferencia del nivel descriptivo, que organiza y cuantifica variables ya establecidas, el exploratorio conserva un carácter cualitativo y flexible. Esta condición lo distingue de los niveles posteriores de naturaleza cuantitativa, situándolo como un espacio de descubrimiento en el que lo previsto y lo imprevisto se integran en la producción de conocimiento.

1.1 LA FENOMENOLOGÍA

La fenomenología es una corriente filosófica y metodológica que estudia la experiencia tal como aparece a la conciencia, sin presupuestos previos. Iniciada por Edmund Husserl y desarrollada por Heidegger, Gadamer, Sartre y Merleau-Ponty, se aplica en ciencias sociales y de la salud para explorar la vivencia subjetiva y los significados que las personas atribuyen a su realidad.

Fenomenología pura o trascendental (Husserl, 1900–1913). Edmund Husserl definió la fenomenología como una “ciencia de las esencias”, orientada a describir rigurosamente las estructuras de la conciencia. Su método central, la reducción fenomenológica (epoché), busca suspender prejuicios para captar los fenómenos tal como aparecen (16). Esta perspectiva no interpreta ni explica, sino que identifica y describe vivencias en su forma más pura, constituyendo el fundamento epistemológico de toda la tradición fenomenológica posterior.

Fenomenología hermenéutica (Heidegger, Gadamer, 1927–1960). Con Martin Heidegger, la fenomenología adquirió un giro hermenéutico. Donde el ser humano habita un mundo de significados, por lo que comprender implica necesariamente interpretar (17). Hans-Georg Gadamer reforzó esta idea con el círculo hermenéutico, donde la comprensión se construye en el diálogo entre las partes y el todo (18). Así, la fenomenología hermenéutica se centra en el sentido atribuido a la experiencia.

Fenomenología existencial (Sartre, Merleau-Ponty, 1940s–1950s). La fenomenología existencial, desarrollada por Jean-Paul Sartre y Maurice Merleau-Ponty, resaltó la existencia concreta y encarnada. Sartre enfatizó la libertad y responsabilidad radical del individuo (19), mientras Merleau-Ponty destacó la percepción y la corporalidad como dimensiones constitutivas (20). Esta corriente abrió camino para comprender cómo los sujetos experimentan la enfermedad, la discapacidad o la muerte, aportando una visión humanista de gran relevancia en salud y ciencias sociales.

1.2. LA HERMENÉUTICA

La hermenéutica es la disciplina dedicada a la interpretación de significados. Nacida en el análisis de textos religiosos y jurídicos, en los siglos XIX y XX se transformó en un enfoque filosófico que busca comprender cómo los seres humanos producen y reinterpretan sentidos en distintos contextos. Más que un método técnico, se consolidó como una teoría general de la comprensión, organizada en tres corrientes principales: la clásica, la filosófica y la crítica.

Hermenéutica clásica (Schleiermacher, Dilthey, siglo XIX). La hermenéutica clásica sistematizó la interpretación como metodología con reglas explícitas. Schleiermacher propuso reconstruir la intención del autor a través del análisis gramatical e histórico (21). Dilthey amplió esta propuesta al sostener que las ciencias del espíritu debían fundamentarse en la comprensión hermenéutica, diferenciándolas de las ciencias naturales (22). En este nivel, comprender equivalía a explicar el sentido original de un texto en su contexto.

Hermenéutica filosófica (Heidegger, Gadamer, 1927–1960). Con Heidegger, la hermenéutica se convirtió en una ontología de la comprensión. El ser humano (Dasein) siempre interpreta desde un horizonte de sentido, nunca de manera neutra (17). Gadamer consolidó esta visión con el círculo hermenéutico y la fusión de horizontes, que muestran la interacción entre tradición y presente (18). Aquí la tarea central deja de ser explicar y pasa a interpretar.

Hermenéutica crítica (Habermas, Ricoeur, 1970–1980s). La hermenéutica crítica añadió una dimensión reflexiva y emancipadora. Habermas sostuvo que todo lenguaje está atravesado por ideología, por lo que interpretar exige también desenmascarar distorsiones (23). Ricoeur complementó esta visión con la hermenéutica de la sospecha, destinada a descubrir supuestos ocultos y a vincular la comprensión con la acción ética y social (24). Su propósito ya no es solo comprender, sino también cuestionar y transformar.

1.3. LA HEURÍSTICA

La heurística es un conjunto de estrategias para el descubrimiento y resolución de problemas en contextos de incertidumbre. Su función es orientar la creatividad, facilitar la toma de decisiones y guiar la investigación cuando no existen soluciones preestablecidas. En ciencia, actúa como recurso exploratorio: ayuda a generar hipótesis, delimitar variables y diseñar caminos iniciales. Así, se convierte en un puente entre la intuición y la formulación metodológica.

Heurística clásica o general. La heurística clásica, vinculada al *ars inveniendi* o “arte de descubrir”, se basa en analogías, metáforas y razonamientos inductivos. Su objetivo es estimular la creatividad y abrir rutas posibles para indagar fenómenos poco comprendidos. No busca respuestas definitivas, sino orientaciones tentativas que faciliten nuevas preguntas. En este nivel, funciona como un instrumento intelectual flexible, generando ideas preliminares que orientan fases posteriores de investigación (25,26).

Heurística metodológica. La heurística metodológica traslada esos recursos a la práctica científica sistemática. Se aplica en la formulación de hipótesis, la selección de variables y el diseño exploratorio de estudios. Aunque mantiene apertura, introduce criterios operativos que aumentan la validez del hallazgo. De este modo, convierte intuiciones iniciales en propuestas metodológicas verificables, siendo el núcleo de la heurística en el nivel exploratorio de la investigación (26,27).

Heurística aplicada o técnica. La heurística aplicada se concreta en reglas prácticas o algoritmos usados cuando no es posible una solución óptima. Destaca en campos como informática, psicología cognitiva o investigación operativa. En salud, puede emplearse para diseñar modelos predictivos preliminares, simplificar diagnósticos iniciales o analizar grandes volúmenes de datos. Así, traduce la heurística en procedimientos operativos, acercándola al análisis cuantitativo y a la práctica profesional (28,29).

1.4. EL CONSTRUCTIVISMO

El constructivismo aplicado a la investigación científica sostiene que el conocimiento no es copia de la realidad, sino una construcción activa que combina teoría, experiencia y cultura. Bajo esta perspectiva, investigar significa producir categorías, modelos y explicaciones situadas, reconociendo el papel del investigador como constructor. Sus vertientes más influyentes son el constructivismo cognitivo, social y radical.

Constructivismo cognitivo (Jean Piaget). El conocimiento se construye mediante asimilación y acomodación, procesos que reorganizan los esquemas mentales a lo largo del desarrollo. En investigación, este enfoque permite diseñar estudios centrados en la evolución cognitiva y en cómo los sujetos generan estructuras cada vez más complejas. Su aporte metodológico radica en analizar los mecanismos internos del aprendizaje como parte del proceso científico. Así, el investigador documenta patrones que explican la construcción progresiva del conocimiento (30).

Constructivismo social (Lev Vygotsky; Berger & Luckmann). El conocimiento se produce en la interacción social, mediado por lenguaje, cultura e instituciones. Vygotsky propuso la zona de desarrollo próximo como concepto clave para explicar el aprendizaje colaborativo (31). Berger y Luckmann ampliaron esta visión mostrando que la realidad misma es una construcción social sustentada en prácticas compartidas (32). En investigación, este enfoque respalda metodologías participativas y cualitativas que exploran cómo los significados se negocian colectivamente.

Constructivismo radical (Ernst von Glasersfeld). El conocimiento no refleja una realidad objetiva, sino que constituye una construcción subjetiva y viable. Para la investigación científica, esto implica reconocer la imposibilidad de acceder a verdades absolutas y valorar la coherencia interna de las teorías como criterio de validez. Se promueve una actitud reflexiva del investigador frente a su propio proceso de construcción. Este enfoque introduce un giro epistemológico fuerte, situando la ciencia como práctica situada y relativa (33).

Cuadro 6. Análisis de datos no estadístico

Enfoque	Definición breve	Ejemplo aplicado
Fenomenología	Busca describir y comprender la experiencia vivida por los sujetos de investigación, resaltando cómo perciben y dan sentido a un fenómeno específico.	Vivencias de pacientes post-COVID-19 en su proceso de rehabilitación domiciliaria, analizadas para identificar percepciones de afrontamiento y recuperación.
Hermenéutica	Procedimiento interpretativo aplicado al análisis de discursos, textos o narrativas, orientado a desentrañar significados, intenciones y contextos en la producción del conocimiento científico.	Discursos presidenciales emitidos durante la pandemia examinados para interpretar estrategias narrativas y sus efectos en la percepción social del riesgo.
Heurística	Promueve la búsqueda de soluciones a partir de la exploración, la creatividad y el aprendizaje derivado de experiencias previas documentadas.	Modelo de educación virtual diseñado para contrarrestar la desinformación sobre COVID-19, tomando como base experiencias previas exitosas de alfabetización digital.
Constructivismo	Asume que el conocimiento se construye colectivamente a través de la interacción social, y que el investigador participa en la configuración de esa construcción.	Prácticas comunitarias frente al contagio en zonas rurales analizadas mediante talleres participativos, orientadas a comprender la construcción social del riesgo.

1.5. EL ESTUDIO HISTÓRICO

Su concepto clásico se fundamenta en la historia como ciencia social orientada a reseñar, analizar y explicar sucesos pasados. Desde esta perspectiva, se asume que el método histórico es esencialmente analítico-sintético: la heurística permite descubrir, recopilar y ordenar fuentes de información, mientras que la hermenéutica posibilita interpretar e integrar esos hallazgos en una explicación más amplia. Así, el análisis se mueve en doble dirección: de lo general a lo particular y de lo particular a lo general, utilizando submétodos cronológicos, geográficos y etnográficos, además de las ciencias auxiliares (arqueología, paleografía, diplomática, entre otras) para garantizar rigor y validez en la reconstrucción de los hechos (34)

Aplicado al ámbito de la salud individual, el estudio histórico se materializa en la historia clínica como relato documentado de la trayectoria de un paciente. Este registro constituye un documento histórico personal, cuya heurística se centra en la búsqueda y sistematización de datos clínicos (síntomas, antecedentes, diagnósticos, intervenciones), mientras que la hermenéutica interpreta esos datos en relación con el contexto vital del sujeto. Por ejemplo, la historia clínica de un adulto con hipertensión arterial no solo consigna cifras tensionales, tratamientos y complicaciones, sino que al contextualizarla revela cómo factores familiares, sociales y culturales han condicionado la evolución de su enfermedad.

En el plano poblacional, el estudio histórico epidemiológico se orienta a la reconstrucción de trayectorias de salud-enfermedad en comunidades. Aquí, la heurística se traduce en la revisión de fuentes documentales, y la hermenéutica en la interpretación de esas series para explicar patrones, rupturas y continuidades en la ocurrencia de enfermedades. Como ejemplo, puede considerarse la investigación de los brotes de cólera en Cuba durante el siglo XIX, donde la revisión cronológica y geográfica de las defunciones permitió comprender cómo las condiciones ambientales, las rutas de navegación y las prácticas culturales influyeron en la propagación del agente infeccioso. (34).

1.6. EL ESTUDIO DOCUMENTAL

El estudio documental es ampliamente utilizado en la investigación científica, especialmente en las ciencias sociales y la salud pública. Este tipo de estudio se basa en la recolección, análisis e interpretación de datos provenientes de fuentes secundarias, como libros, artículos científicos, informes técnicos, bases de datos y otros documentos relevantes para el área de interés. Las fuentes utilizadas deben ser fiables, pertinentes y actualizadas (35).

Este tipo de estudio requiere una organización rigurosa para clasificar, categorizar y sintetizar los datos extraídos de los documentos. En la salud pública los estudios documentales son esenciales para identificar patrones, tendencias y antecedentes históricos de enfermedades, intervenciones y políticas. Aunque es común como fase inicial en investigaciones más amplias, el estudio documental también puede ser una metodología completa para responder a preguntas específicas de investigación.

En el ámbito de la salud pública (35), el estudio documental es una herramienta invaluable para: Revisar literatura: Proporciona un marco teórico sólido para investigaciones y permite contextualizar problemas de salud dentro de un panorama más amplio. Evaluación de políticas: Ayuda a analizar documentos normativos y reglamentarios para identificar fortalezas y áreas de mejora en programas de salud. Generación de hipótesis: Un estudio documental debe plantear claramente su objetivo.

Las fuentes deben seleccionarse con criterios de relevancia, confiabilidad y actualidad (14). Los datos recolectados se clasifican en categorías o temas específicos. Se utilizan herramientas cualitativas para interpretar los hallazgos. Los resultados se presentan de forma estructurada y enfocada en responder a las preguntas de investigación. Algunas limitaciones que deben ser consideradas, son la dependencia de información existente. En algunos casos, los documentos pueden no reflejar las tendencias más recientes. La selección e interpretación de documentos puede estar influenciada por sesgos del investigador (36).

1.7. EL ESTUDIO DEL DISCURSO

El estudio del discurso es un campo interdisciplinario que examina el uso del lenguaje en contextos específicos, considerando factores lingüísticos, sociales, culturales y psicológicos (37). El discurso puede definirse como una forma de comunicación que trasciende las palabras y frases, involucrando un contexto más amplio que incluye a los participantes, sus propósitos y las circunstancias en que se desarrolla (38).

En este sentido, el estudio del discurso se centra en textos escritos, conversaciones, discursos políticos, mensajes mediáticos y cualquier forma de comunicación humana. En el contexto de la clínica y la salud pública, el discurso es fundamental para comprender cómo se establece la comunicación entre el profesional de la salud y el paciente, especialmente durante la anamnesis. Este proceso de recolección de información médica no solo identifica signos y síntomas, sino que también explora las percepciones del paciente sobre su salud, permitiendo una visión integral que considera los determinantes sociales y culturales (39).

Desde una perspectiva teórica, el estudio del discurso abarca varias áreas. La pragmática, estudia el significado en contexto, y el análisis de la conversación se centra en cómo interactúan las personas en tiempo real. En el ámbito clínico, facilita una anamnesis efectiva, ayudando al profesional de la salud a establecer diagnósticos más precisos y diseñar intervenciones adaptadas a las necesidades del paciente.

El estudio del discurso utiliza metodologías cualitativas, como el análisis en profundidad de textos y conversaciones, el uso de herramientas tecnológicas para procesar grandes cantidades de datos textuales y la observación etnográfica que combina el discurso con su contexto cultural y social (39). El estudio del discurso es esencial para entender cómo el lenguaje influye en la sociedad. En la salud pública y la clínica, su aplicación optimiza la comunicación entre profesionales de la salud y pacientes, mejorando procesos como la anamnesis (37).

1.8. EL ESTUDIO ETNOGRÁFICO

Es una metodología cualitativa que busca comprender los fenómenos sociales desde la perspectiva de los participantes, explorando sus comportamientos, costumbres, valores y significados culturales (40). Este enfoque se fundamenta en la observación directa y la inmersión prolongada del investigador en el contexto estudiado, con el fin de obtener una comprensión profunda y contextualizada de las realidades humanas.

La etnografía, originada en la antropología, ha evolucionado para aplicarse a diversos campos, incluyendo la salud pública (41). En este ámbito, resulta crucial para explorar prácticas de salud, creencias sobre enfermedades y los determinantes culturales que influyen en el comportamiento relacionado con la salud. Su enfoque en los contextos sociales y culturales es esencial para diseñar intervenciones que se adapten a las necesidades específicas de las comunidades. Permite comprender cómo las comunidades perciben la salud y la enfermedad, lo que facilita la identificación de barreras culturales que puedan obstaculizar la aceptación de programas como la vacunación (42). Además, contribuye al diseño de estrategias de intervención que respeten las prácticas y valores locales, promoviendo una participación de las comunidades en la construcción de soluciones sostenibles.

El estudio etnográfico emplea diversas técnicas para la recolección de datos, entre ellas se destacan la observación participante, donde el investigador se integra en la comunidad y participa en sus actividades cotidianas (43); las entrevistas abiertas, que permiten recoger narrativas personales y perspectivas únicas; y la revisión de documentos locales, como textos o artefactos culturales.

Algunos desafíos de esta metodología, incluyen que la interpretación de los datos puede estar influenciada por los sesgos del investigador (42). Por otro lado, es un proceso intensivo en tiempo, que requiere un compromiso prolongado para garantizar una inmersión adecuada. Además, sus hallazgos suelen ser específicos del contexto estudiado, lo que limita su generalización a otras comunidades.

1.9. EL ESTUDIO DE CASO

El estudio de caso se concibe clásicamente como un método de investigación cualitativa que permite analizar en profundidad una unidad específica, ya sea un individuo, un grupo, una institución o un proceso social, con el propósito de comprender sus particularidades y, a partir de ellas, generar conocimiento transferible a contextos semejantes. Este enfoque se caracteriza por su capacidad de integrar distintas técnicas de recolección de información (observación, entrevistas, documentos, registros) y por privilegiar la perspectiva holística (44). El estudio de caso, en consecuencia, se constituye como una herramienta heurística que permite describir, interpretar y, en algunos casos, explicar fenómenos sociales o de salud en escalas pequeñas, sin pretensión de generalización estadística, pero con valor analítico y conceptual.

En el ámbito clínico, el estudio de caso se plasma en la narración de la trayectoria de un paciente, integrando tanto el proceso diagnóstico y terapéutico como sus condiciones sociales y personales. Por ejemplo, puede considerarse el estudio de caso de una mujer de 45 años con diagnóstico de cáncer cervicouterino. El valor metodológico de este estudio de caso reside en la integración de los datos biomédicos con la interpretación de factores culturales y conductuales (como el acceso irregular a tamizajes previos), lo que permite comprender la enfermedad en un marco más amplio que el estrictamente biológico (44).

A nivel epidemiológico, el estudio de caso se orienta al análisis de un evento de salud en una comunidad o institución, con el fin de identificar causas, dinámicas y posibles medidas de intervención. Un ejemplo ilustrativo sería el estudio de caso de un brote de influenza en una escuela secundaria: el análisis documenta la aparición de los primeros contagios, la velocidad de propagación en las aulas, los factores ambientales (ventilación, hacinamiento), las prácticas culturales (compartir alimentos, escaso uso de mascarillas) y la respuesta institucional. Este estudio no pretende generalizar a toda la población escolar del país, pero sí generar hipótesis y aprendizajes útiles para diseñar protocolos de prevención y control en contextos semejantes (44).

1.10. LA TEORÍA FUNDAMENTADA

La teoría fundamentada (Grounded Theory), desde su formulación original en 1967 hasta las versiones contemporáneas, ha transitado por distintos momentos que no solo marcan una secuencia temporal, sino que representan enfoques metodológicos diferenciados, cada uno con su propia postura epistemológica y aporte distintivo.

El enfoque clásico inductivo, desarrollado por Glaser y Strauss (1967), representa el nacimiento del método. Su propuesta se sitúa dentro de un positivismo moderado, donde el investigador busca mantener una posición de neutralidad frente al fenómeno. La teoría, emerge directamente de los datos a través de procedimientos inductivos como la codificación abierta y la comparación constante, sin partir de hipótesis previas (45).

Posteriormente, con Strauss y Corbin (1990), surge el enfoque sistematizado, que implica una formalización técnica del proceso analítico. Se enmarca en el postpositivismo estructurado, y se caracteriza por la introducción de fases bien definidas de codificación: abierta, axial y selectiva. Este modelo busca aumentar la rigurosidad procedimental y facilitar la enseñanza del método a investigadores en formación (46).

El enfoque constructivista, de Kathy Charmaz (2006), supone un cambio de paradigma respecto a las versiones anteriores. Se posiciona en el constructivismo interpretativo, reconociendo que el conocimiento no se descubre, sino que se co-construye entre el investigador y los participantes. En este modelo, se acepta la subjetividad, la reflexividad y el papel del contexto en la generación del conocimiento (47).

Finalmente, Adele Clarke (2005) plantea un enfoque posmoderno, que implica una expansión crítica del método original. Desde una postura posmoderna y situacional, propone herramientas como mapas analíticos, mapas de posicionamiento y análisis de discursos, integrando relaciones de poder, actores no humanos y complejidades contextuales. Este enfoque permite abordar fenómenos sociales desde una perspectiva más amplia y crítica, superando los límites del análisis lineal tradicional (48).

Cuadro 7. Diseños de investigación cualitativa

Nivel de profundidad	Tipo de estudio	Propósito principal	Técnicas utilizadas
~ Bajo	Estudio documental	Analizar fuentes secundarias para identificar patrones, antecedentes y generar hipótesis	Revisión de literatura, análisis temático, categorización, síntesis crítica
~ Medio	Estudio histórico	Comprender la evolución de fenómenos de salud a lo largo del tiempo	Análisis de documentos históricos, informes, registros médicos antiguos
~ Medio	Estudio del discurso	Examinar el lenguaje y su influencia en la práctica clínica y la salud pública	Análisis textual, conversación, observación etnográfica, herramientas computacionales
~ Alto	Estudio etnográfico	Entender prácticas y significados culturales desde la perspectiva interna de los actores	Observación participante, entrevistas abiertas, revisión de documentos culturales
~ Alto	Estudio de caso	Analizar un fenómeno en profundidad para obtener aprendizajes transferibles	Entrevistas, revisión documental, análisis clínico y observacional intensivo
~ Muy alto	Teoría fundamentada	Generar teorías fundamentadas en los datos recolectados mediante análisis inductivo	Codificación abierta, memos analíticos, comparación constante, muestreo teórico

1.11. OBJETIVOS DEL NIVEL EXPLORATORIO

En el nivel exploratorio, los objetivos no se formulan como en los estudios cuantitativos, donde se plantean en términos medibles y verificables. Aquí los propósitos cumplen una función de guía metodológica, orientando el proceso de descubrimiento y delimitación del fenómeno. Su papel es organizar la búsqueda de información y preparar la transición hacia fases más rigurosas, sin pretender establecer resultados concluyentes.

El primer propósito es identificar (reconocer) fenómenos, variables o dimensiones emergentes vinculadas al objeto de estudio. La tarea consiste en registrar los elementos aparezcan tal como se presentan en la realidad. Se utilizan herramientas como revisión documental, observación abierta y entrevistas preliminares. El producto es un inventario inicial de aspectos relevantes para la investigación.

El segundo propósito busca interpretar (comprender) los significados que los actores atribuyen a sus experiencias y prácticas. Se apoya en enfoques fenomenológicos y hermenéuticos, con énfasis en el contexto social y cultural. No pretende generalizar, sino captar la riqueza subjetiva y situada de los discursos. El resultado es una comprensión profunda que permite avanzar hacia marcos conceptuales preliminares.

El tercer propósito consiste en determinar (precisar) los aspectos centrales del problema y los vacíos de conocimiento. Supone pasar de un registro amplio a la focalización en factores prioritarios para la investigación. Se emplean estrategias como triangulación de fuentes, matrices heurísticas y comparación de evidencias. El producto es la delimitación del núcleo problemático, que orientará el diseño posterior.

El cuarto propósito es definir (operacionalizar) las categorías preliminares en criterios observables y medibles. Esta etapa prepara la transición hacia el nivel descriptivo, consolidando glosarios, unidades de análisis e indicadores tentativos. No busca cerrar el proceso exploratorio, sino formalizarlo metodológicamente. El resultado es un marco operativo inicial que conecta lo cualitativo con lo cuantitativo.

1.12. OBJETIVO IDENTIFICAR (RECONOCER)

En el nivel exploratorio, el objetivo es reconocer fenómenos o variables emergentes mediante la revisión documental, entrevistas abiertas y observación de campo, generando un inventario preliminar que sirva como base para la investigación posterior.

En el nivel exploratorio, el análisis de datos asociado al objetivo identificar (reconocer) es de carácter abierto, inductivo y descriptivo. No busca confirmar hipótesis ni probar relaciones causales, sino descubrir fenómenos emergentes a partir de la realidad observada. Su finalidad es mantener la flexibilidad metodológica necesaria para captar lo nuevo, lo inesperado o lo poco estudiado en el objeto de investigación.

Las técnicas principales incluyen la revisión documental inicial, la observación de campo y las entrevistas abiertas. También se emplea la codificación preliminar, que permite organizar fragmentos de información en categorías tentativas. Estas técnicas funcionan como herramientas de exploración que facilitan un acercamiento integral, captando tanto datos objetivos como percepciones y narrativas de los participantes.

El producto esperado es un inventario preliminar de categorías, dimensiones o variables emergentes. Este inventario no es definitivo, pero constituye la primera aproximación ordenada al fenómeno, ayudando a delimitar el problema y a plantear nuevas preguntas de investigación. En esta etapa, los hallazgos se expresan en listados, matrices o esquemas conceptuales que organizan la información dispersa.

La importancia del análisis de datos en este objetivo radica en que establece la base metodológica para los niveles posteriores. Al identificar fenómenos emergentes, se sientan las condiciones para su interpretación, determinación y posterior definición operativa. Así, el objetivo identificar (reconocer) actúa como puente inicial entre la observación del fenómeno y la construcción sistemática del conocimiento científico.

1.13. OBJETIVO INTERPRETAR (COMPRENDER)

El segundo propósito es interpretar los significados que los sujetos atribuyen a sus experiencias, mostrando cómo construyen sentido frente al fenómeno. Mediante análisis narrativo y temático, a partir de entrevistas y testimonios, se logra una comprensión contextual y profunda de las vivencias en sus dimensiones culturales y sociales.

El análisis de datos orientado al objetivo interpretar (comprender) es eminentemente cualitativo, hermenéutico y contextual. Busca captar los significados y sentidos que los sujetos atribuyen a sus experiencias y prácticas. En esta etapa, los datos no se limitan a enumerar hechos, sino que se procesan como testimonios que requieren comprensión en su marco cultural e histórico.

Las técnicas empleadas incluyen el análisis narrativo, el análisis temático y la construcción de matrices de significados. También se utiliza el círculo hermenéutico, que permite ir y venir entre las partes y el todo, logrando una comprensión más completa. Estas herramientas posibilitan que el investigador se acerque al trasfondo simbólico y social de los discursos recogidos.

El producto esperado en este objetivo es una comprensión profunda de significados y sentidos construidos por los actores sociales. Este resultado proporciona un mapa interpretativo del fenómeno, que visibiliza tanto dimensiones explícitas como implícitas de la experiencia. En muchos casos, permite descubrir categorías analíticas que no eran evidentes en la fase inicial de identificación.

El análisis de datos en este objetivo aporta un nivel de densidad conceptual que enriquece la investigación. Interpretar no solo abre el fenómeno a múltiples lecturas, sino que permite construir vínculos teóricos preliminares con marcos de referencia existentes. Así, la etapa de interpretación conecta el descubrimiento inicial con una primera sistematización conceptual.

1.14. OBJETIVO DETERMINAR (PRECISAR)

El tercer objetivo busca determinar (precisar) los aspectos centrales del problema, delimitando vacíos de conocimiento, factores prioritarios y líneas de indagación. Se trata de pasar de un registro abierto a una selección más focalizada de categorías y dimensiones clave. Para ello, se emplean técnicas de comparación de fuentes, triangulación de datos y matrices heurísticas. El producto esperado es una definición clara del núcleo problemático, que oriente la construcción de objetivos en fases posteriores.

En el nivel exploratorio, el análisis de datos para determinar (precisar) se caracteriza por ser comparativo y selectivo, de naturaleza heurística. Su propósito es pasar de la amplitud inicial a la delimitación de los aspectos centrales del problema. Esto permite focalizar el análisis en factores clave y dejar de lado los elementos secundarios.

Las técnicas principales son la triangulación de fuentes, las matrices heurísticas y el conteo cualitativo de menciones. Estas herramientas facilitan establecer qué categorías o dimensiones aparecen con mayor frecuencia o relevancia en las fuentes revisadas. El énfasis está en contrastar y priorizar, más que en generar nuevos hallazgos.

El producto esperado es la delimitación del núcleo problemático, es decir, un esquema de los factores prioritarios y los vacíos de conocimiento detectados. Esta delimitación permite depurar el inventario inicial y seleccionar las dimensiones que tendrán continuidad en fases posteriores.

El valor de este análisis radica en que aporta claridad y foco metodológico. Determinar y precisar no solo organiza el fenómeno, sino que prepara las condiciones para avanzar hacia definiciones más formales. Es un puente entre la apertura cualitativa y la preparación para la cuantificación, consolidando la estructura de la investigación.

1.15. OBJETIVO DEFINIR (OPERACIONALIZAR)

El nivel exploratorio culmina con el objetivo de definir (operacionalizar) los conceptos emergentes, transformándolos en criterios observables y medibles. Este paso traduce las categorías cualitativas en variables tentativas, indicadores y unidades de análisis, preparando la transición hacia el nivel descriptivo. Se utilizan glosarios conceptuales, esquemas de variables e instrumentos piloto. El producto esperado es un marco operativo inicial, que permitirá aplicar técnicas estadísticas en fases posteriores de la investigación.

El análisis de datos para definir (operacionalizar) tiene un carácter de categorización formal, bajo un enfoque constructivista. Se centra en transformar conceptos emergentes en criterios observables y medibles, que sirvan como base para estudios posteriores. Aquí, el investigador comienza a preparar el terreno para la estadística descriptiva.

Las técnicas utilizadas incluyen la construcción de definiciones operativas, la elaboración de glosarios conceptuales y el diseño de indicadores tentativos. Estas herramientas traducen las categorías cualitativas en variables de análisis, con criterios explícitos de observación y medición.

El producto esperado es un marco operativo inicial, que organiza las dimensiones del fenómeno en variables observables. Este marco constituye un puente metodológico entre el nivel exploratorio y el descriptivo, permitiendo la aplicación de instrumentos cuantitativos en fases futuras.

La importancia de este análisis radica en que consolida la progresión metodológica del nivel exploratorio. Definir operativamente no significa cerrar el proceso, sino preparar la transición hacia la medición estadística. De este modo, se logra continuidad entre la exploración cualitativa inicial y la descripción cuantitativa posterior.

NIVEL

DESCRIPTIVO

ANÁLISIS ESTADÍSTICO UNIVARIADO

Los estudios de nivel descriptivo son investigaciones cuantitativas que tienen como finalidad caracterizar de manera sistemática las propiedades de una o más variables de una población o muestra. Describe cómo se presenta el fenómeno en la realidad observada, permitiendo construir una línea base de conocimiento que documente con precisión el estado del objeto de estudio en un contexto y momento determinados.

Este nivel se diferencia del exploratorio porque no busca descubrir o definir el fenómeno, sino cuantificar variables previamente identificadas; además, utiliza estadística en lugar de métodos inductivos no estadísticos como la observación o la codificación abierta.

A su vez, se distingue del nivel relacional, que ya introduce análisis bivariado con el objetivo de establecer asociaciones o correlaciones entre variables; en cambio, el nivel descriptivo no compara ni vincula variables entre sí, sino que se limita a describirlas individualmente dentro de la muestra o población estudiada.

2.1. LA ESTADÍSTICA EN LA INVESTIGACIÓN CIENTÍFICA

La estadística constituye la base cuantitativa para transformar datos en evidencia. Su uso permite describir, analizar y comparar fenómenos, asegurando que las conclusiones se sustenten en medidas objetivas y reproducibles. La conexión inicial se establece con el análisis estadístico univariado. Este primer nivel de análisis es la herramienta fundamental para el nivel descriptivo de la investigación, y representa el punto de partida para procedimientos más complejos como el análisis bivariado y multivariado. La estadística se articula en dos grandes ramas: la descriptiva y la inferencial.

La estadística descriptiva, como señalan Posada Hernández (49) y Mas (50), se centra en sintetizar la información obtenida de una muestra o población mediante medidas de tendencia central, dispersión y representación gráfica, sin extraer conclusiones más allá de los datos observados. Es especialmente útil en investigaciones de carácter univariado o en la primera fase de estudios más complejos, pues proporciona el panorama inicial sobre las variables de interés.

La estadística inferencial, constituye el pilar de los estudios que buscan generalizar resultados y realizar estimaciones o contrastes de hipótesis. De acuerdo con Montanero Fernández (51) y Paliz Sánchez (52), esta rama utiliza el cálculo de probabilidades y la teoría muestral para deducir conclusiones sobre una población a partir de una muestra representativa. En investigaciones bivariadas o de nivel relacional, tal como describe Espinoza Pajuelo (53), la estadística adquiere un papel decisivo al permitir evaluar el vínculo y la fuerza de asociación entre dos variables. En el ámbito de la epidemiología y las ciencias de la salud, las medidas de asociación como el RR y el OR (54) son esenciales para estimar la magnitud del vínculo entre factores de exposición y eventos de salud.

El análisis de datos estadístico no se limita a la aplicación de fórmulas o programas de análisis, sino que implica un proceso reflexivo y crítico que abarca desde la planificación del estudio hasta la interpretación final de los resultados.

2.2. ESTUDIOS NO INFERENCIALES

Se caracteriza por el análisis de los datos sin intención de extrapolar los resultados obtenidos en una muestra hacia la población total ni de someter proposiciones a contrastes estadísticos de hipótesis. En consecuencia, su objetivo central es describir y organizar la información recolectada, permitiendo identificar las características y patrones que presenta el fenómeno de interés en el grupo estudiado (49).

En la clasificación general de los niveles de investigación cuantitativa, los estudios no inferenciales se ubican dentro de la estadística descriptiva, la cual se ocupa de resumir, presentar y analizar un conjunto de datos sin aplicar procedimientos de estimación o pruebas de significancia (55). Estos estudios se distinguen de los inferenciales en que no establecen conclusiones probabilísticas ni determinan márgenes de error, centrándose, en cambio, en la observación sistemática y la presentación estructurada de la información.

La interpretación de resultados se limita al grupo o contexto estudiado, sin pretender su extrapolación a otras poblaciones o entornos (49). Ausencia de prueba de hipótesis: no se aplican contrastes estadísticos para aceptar o rechazar proposiciones formales, lo que evita la dependencia de supuestos probabilísticos propios de la estadística inferencial.

Etapas de caracterización, la intención de un estudio no es inferencial de modo que, cuando estamos frente a una variable numérica sus descriptivos son las medidas de tendencia central (media, mediana y moda) y las medidas de dispersión (desviación estándar, varianza y error típico de la media). Mientras que estando frente a una variable categórica sus descriptivos serán las frecuencias absolutas y las frecuencias relativas, ya sean acumuladas o no acumuladas, lo cual depende de la presentación de resultados (56). Las gráficas incluyen a los histogramas y barras de error en caso de variables numéricas; y diagramas de sectores y gráfico de barras para los casos de variables categóricas.

2.3. ESTUDIOS INFERENCIALES

Los estudios inferenciales permiten proyectar los resultados obtenidos en una muestra hacia la población de la cual esta procede, siempre considerando un margen de error y un nivel de confianza previamente definidos. A diferencia de los estudios no inferenciales, que se limitan a describir y organizar los datos, los estudios inferenciales utilizan técnicas estadísticas que posibilitan la generalización de conclusiones y, en muchos casos, la contrastación de hipótesis acerca de parámetros poblacionales desconocidos.

Etapas de estimación de parámetros, estamos hablando claramente de una intención inferencial (55), cuando la variable es numérica, el estimador de la media poblacional es el promedio de la muestra estudiada y sus intervalos de confianza se calculan a partir del error estándar de la media como medida de variabilidad, habitualmente con un nivel de confianza del 95% (55). Cuando el parámetro a estimar es un valor categórico, como la prevalencia de una enfermedad en la población la variable categórica dicotómica como presencia (p) y ausencia (q), estimando el valor poblacional de p con sus respectivos intervalos de confianza habitualmente al 95%.

Etapas de contraste con un parámetro, corresponde la prueba de hipótesis, dado (55). En el caso de las variables numéricas se plantea la siguiente hipótesis: H_0 : El valor de la media del grupo no es distinto al parámetro de la población, H_1 : El valor de la media del grupo es distinta al parámetro de la población. Habitualmente con un nivel de significancia del 5% y la prueba de t de Student para una sola muestra, con la cual se calcula el valor de p, a fin de concluir con una de las hipótesis.

En el caso de las variables categóricas, la hipótesis sería: H_0 : El valor de la proporción grupo no es distinto a la prevalencia de la población, H_1 : El valor de la proporción del grupo es distinta a la prevalencia de la población, habitualmente con un nivel de significancia del 5% y la prueba de χ^2 (ji-cuadrado) de bondad de ajuste, con la cual se calcula el valor de p, a fin de concluir con una de las hipótesis (55).

Cuadro 8: Etapas del análisis estadístico

Análisis estadístico	Etapas	Propósito principal	Técnicas estadísticas empleadas
No inferencial	Caracterización	Describir variables de interés y de caracterización de manera individual, sin generalizar los resultados a la población.	<ul style="list-style-type: none"> - Variables numéricas: media, mediana, moda, desviación estándar, varianza, rango, percentiles. - Variables categóricas: frecuencias absolutas, relativas y acumuladas. - Representaciones gráficas: histogramas, diagramas de barras, diagramas de sectores.
Inferencial	Estimación de parámetros	Calcular un valor aproximado del parámetro poblacional (media o proporción) a partir de una muestra representativa, acompañado de un intervalo de confianza, sin realizar pruebas de hipótesis.	<ul style="list-style-type: none"> - Estimación puntual (media o proporción). - Intervalo de confianza, habitualmente al 95 %. - Media: cálculo usando error estándar. - Proporción: cálculo del margen de error con base en la variabilidad muestral.
Inferencial	Contraste con un parámetro	Evaluar si el valor obtenido en la muestra difiere significativamente del parámetro poblacional conocido, mediante una prueba de hipótesis.	<ul style="list-style-type: none"> - Variables numéricas: prueba t de Student para una muestra. - Variables categóricas: prueba χ^2 (ji-cuadrado) de bondad de ajuste. - Planteamiento de hipótesis nula y alterna. - Nivel de significancia habitual: 5 % ($p < 0,05$).

2.4. OBJETIVO DESCRIBIR O CARACTERIZAR

En el marco de una investigación científica, la correcta identificación y clasificación de las variables resulta esencial para la planificación, el análisis y la interpretación de los resultados. Según Paliz Sánchez (52) y Montanero Fernández (51), la variable de interés (variable de estudio) es aquella que se analiza de forma prioritaria para responder al objetivo principal de la investigación. Representa el fenómeno central que se busca describir. Puede ser cuantitativa si se mide numéricamente, como la presión arterial, la concentración de glucosa o el puntaje en un test, o cualitativa si describe categorías o atributos, como el tipo de tratamiento, el grupo sanguíneo o el diagnóstico clínico.

En función al nivel investigativo, la variable de interés puede adoptar diferentes roles, por lo que su definición debe ser precisa, operacionalizable y coherente con el alcance del estudio, incluyendo el nivel de medición (nominal, ordinal, de intervalo o de razón) a fin de ajustar la técnica de recolección de datos.

Por otro lado, las variables de caracterización (variables auxiliares) corresponden a factores que no forman parte directa de la pregunta principal de investigación, pero que describen el contexto de la población o de los sujetos estudiados. Como indican Posada Hernández (49) y Espinoza Pajuelo (53), estas variables permiten comprender mejor las condiciones bajo las cuales se obtienen los resultados y facilitan la segmentación, y comparación. Entre ellas se incluyen características sociodemográficas (edad, sexo, nivel educativo), clínicas (antecedentes médicos, comorbilidades), geográficas (lugar de residencia, zona urbana o rural) o institucionales (tipo de centro de salud, disponibilidad de recursos).

La inclusión de las variables de caracterización en la base de datos y su análisis descriptivo aportan valor al estudio, ya que ayudan a interpretar si los resultados observados en la variable de interés son consistentes en distintos subgrupos o si pueden estar influenciados por condiciones particulares de la muestra. El análisis estadístico de los datos dependerá de la naturaleza de las variables.

El análisis descriptivo de una variable numérica implica resumir la información contenida en los datos mediante indicadores estadísticos. Tal como exponen Paliz Sánchez (52), Posada Hernández (49) y Mas (50), este proceso abarca cuatro grandes grupos de medidas: tendencia central, dispersión, posición y forma.

Las medidas de tendencia central buscan representar un valor típico o representativo de la distribución. La media aritmética es la más utilizada, calculada como la suma de todos los valores dividida entre el número de observaciones, y resulta útil cuando la distribución es simétrica y sin valores extremos. La mediana, en cambio, divide la serie ordenada en dos partes iguales y es más robusta ante valores atípicos. La moda corresponde al valor más frecuente en la distribución.

Las medidas de dispersión cuantifican el grado en que los valores se alejan del centro de la distribución. El rango expresa la diferencia entre el valor máximo y el mínimo, siendo sensible a valores atípicos. La varianza mide la desviación media cuadrática respecto de la media, y su raíz cuadrada, la desviación estándar, indica la dispersión en las mismas unidades que la variable original. El coeficiente de variación, permite comparar la variabilidad relativa entre series con diferentes unidades o magnitudes.

Las medidas de posición, localiza un valor dentro de la distribución. Los percentiles dividen los datos ordenados en cien partes iguales, mientras que los cuartiles lo hacen en cuatro y los deciles en diez. Estos indicadores permiten identificar el valor por debajo del cual se sitúa un determinado porcentaje de observaciones, siendo útiles para establecer puntos de corte, clasificaciones o comparaciones entre subgrupos.

Las medidas de forma describen características de la distribución. La asimetría indica el grado y dirección de la falta de simetría: una distribución con asimetría positiva presenta una cola más larga a la derecha, mientras que una asimetría negativa la tiene a la izquierda. La curtosis evalúa el grado de concentración de los valores alrededor de la media; distribuciones leptocúrticas presentan picos más pronunciados, mientras que las platicúrticas muestran formas más achatadas respecto de la distribución normal.

El análisis descriptivo de una variable categórica tiene como finalidad sintetizar y organizar la información recogida en categorías, permitiendo identificar patrones de distribución y comparaciones entre grupos. Según lo expuesto por Posada Hernández (49), Montanero Fernández (51) y Paliz Sánchez (52), este análisis se basa principalmente en el cálculo de frecuencias, que pueden expresarse en forma absoluta, relativa y acumulada, dependiendo de la naturaleza de la variable.

La frecuencia absoluta (n) corresponde al número de observaciones que se registran en cada categoría de la variable. Es la forma más directa de contabilizar los datos y constituye la base para el cálculo de otras medidas. Por ejemplo, en una variable como “grupo sanguíneo” con categorías A, B, AB y O, la frecuencia absoluta indica cuántos individuos pertenecen a cada grupo sin hacer referencia al tamaño total de la muestra.

La frecuencia relativa (%) expresa la proporción que representa cada categoría respecto al total de observaciones. Se obtiene dividiendo la frecuencia absoluta de cada categoría entre el número total de casos y multiplicando por cien. Este valor facilita la comparación entre categorías cuando los tamaños muestrales difieren o cuando se desea interpretar la importancia relativa de cada grupo. Siguiendo el ejemplo anterior, conocer que el 45 % de los participantes tiene grupo sanguíneo O, ofrece una visión más contextualizada que el dato aislado de la frecuencia absoluta.

La frecuencia acumulada se utiliza especialmente en variables categóricas nominales y en variables nominales cuando están ordenadas de mayor a menor y las variables que están naturalmente ordenadas, es decir, aquellas cuyas categorías mantienen un orden natural (por ejemplo, “nivel educativo”: primaria, secundaria, superior). Se obtiene sumando progresivamente las frecuencias absolutas o relativas desde la primera categoría hasta la última, siguiendo el orden establecido. Este indicador permite identificar el porcentaje o número de casos que se encuentran en una categoría determinada o por debajo de ella, lo que resulta útil para describir la distribución acumulativa de la variable.

Buenas prácticas para la presentación de resultados descriptivos

Para garantizar la transparencia y uniformidad en la presentación de resultados descriptivos, se recomienda seguir las guías internacionales de reporte:

- **CONSORT 2010:** para ensayos clínicos, incluye el formato y contenido sugerido para la tabla de características basales (57).
- **STROBE 2007:** para estudios observacionales, especifica cómo presentar las variables descriptivas, indicando medidas de tendencia central, dispersión y proporciones según la naturaleza de las variables (58).

Ejemplo de tabla descriptiva (adaptado de STROBE):

ej. Tabla. Características basales de los participantes.

Variable	Grupo A (n=50)	Grupo B (n=50)	Total (n=100)
Edad, media (DE)	45,2 (10,3)	44,8 (9,7)	45,0 (10,0)
Sexo, n (%)			
- Femenino	28 (56%)	30 (60%)	58 (58%)
- Masculino	22 (44%)	20 (40%)	42 (42%)
IMC, media (DE)	26,5 (3,4)	27,0 (3,6)	26,8 (3,5)

DE: desviación estándar; IMC: índice de masa corporal.

Notas:

- Usar **media (DE)** para variables numéricas.
- Usar **n (%)** para variables categóricas.
- Indicar siempre el tamaño muestral utilizado en cada cálculo si hay datos faltantes.

El análisis gráfico de una variable numérica permite visualizar de forma inmediata la distribución de los datos, identificar patrones, detectar valores atípicos y evaluar la forma de la distribución. De acuerdo con lo expuesto por Posada Hernández (49), Mas (50) y Paliz Sánchez (52) existen diversas representaciones, entre las cuales destacan el histograma, el diagrama de tallo y hojas y el diagrama de caja y bigotes.

El histograma (para conjuntos grandes) es una de las herramientas más empleadas para describir variables numéricas continuas y en ocasiones discretas con numerosos valores. Consiste en dividir el rango total de los datos en intervalos o clases y representar la frecuencia de observaciones en cada uno mediante barras contiguas. La altura de cada barra refleja la frecuencia absoluta o relativa del intervalo, lo que permite apreciar la forma de la distribución (simétrica, asimétrica, unimodal o multimodal), su grado de dispersión y la presencia de concentraciones o vacíos de datos.

El diagrama de tallo y hojas (para conjuntos pequeños) combina elementos gráficos y tabulares, ofreciendo la ventaja de conservar los datos originales al tiempo que muestra su distribución. Se construye dividiendo cada valor en dos partes: el “tallo”, que corresponde a las cifras más significativas (por ejemplo, las decenas), y la “hoja”, que representa las cifras menos significativas (por ejemplo, las unidades). Las hojas se colocan ordenadas junto a su tallo, formando una visualización que permite identificar la forma de la distribución, detectar valores atípicos y observar con precisión cada dato.

El diagrama de caja y bigotes (boxplot) es una representación sintética que resume la distribución de la variable a partir de sus medidas de posición: la mediana, el primer y el tercer cuartil. La “caja” abarca el rango intercuartílico, es decir, el 50 % central de los datos, mientras que los “bigotes” se extienden hasta los valores más extremos que no se consideran atípicos. Los puntos situados fuera de este rango se marcan de forma individual como valores atípicos. Este gráfico permite comparar de manera visual la dispersión, la simetría y la presencia de valores extremos, siendo muy útil tanto en análisis exploratorios como en comparaciones entre grupos.

El análisis gráfico de una variable numérica constituye una herramienta clave para sintetizar y comunicar visualmente la distribución de los datos, facilitando la identificación de patrones y la comparación entre categorías. De acuerdo con lo descrito por Posada Hernández (49), Montanero Fernández (51) y Paliz Sánchez (52), tres de las representaciones más empleadas para este fin son el diagrama de barras, el diagrama de sectores y el diagrama de Pareto.

El diagrama de barras se compone de barras separadas, cuya altura o longitud es proporcional a la frecuencia, ya sea absoluta o relativa, de cada categoría de la variable. Las categorías se sitúan en el eje horizontal, mientras que el eje vertical representa las frecuencias. Este tipo de gráfico permite apreciar de forma inmediata las diferencias en magnitud entre grupos y es aplicable tanto a variables cualitativas nominales (sin orden natural) como ordinales (con jerarquía). Una de sus ventajas es la flexibilidad para ordenar las categorías de forma alfabética, por tamaño de frecuencia o por relevancia para el estudio, lo que facilita un análisis más dirigido.

El diagrama de sectores, o gráfico circular, muestra las proporciones de cada categoría en forma de sectores dentro de un círculo. Cada sector ocupa un ángulo cuya medida es proporcional a la frecuencia relativa de la categoría correspondiente, de manera que el conjunto de sectores suma 360 grados. Esta representación es especialmente útil para destacar la participación porcentual de cada grupo en el total y para comunicar de forma intuitiva la composición de una variable.

El diagrama de Pareto combina el diagrama de barras con una línea que representa las frecuencias acumuladas. Las barras se disponen de mayor a menor frecuencia, mientras que la línea ascendente muestra la acumulación porcentual de las categorías. Este gráfico se basa en el principio de Pareto o regla 80/20, según el cual un número reducido de categorías suele concentrar la mayor parte de la frecuencia total. Su principal utilidad radica en ayudar a priorizar la atención en aquellas categorías que generan el mayor impacto en el fenómeno estudiado. Es muy utilizado en control de calidad, análisis de problemas y toma de decisiones.

2.5. OBJETIVO ESTIMAR O CALCULAR UN PARÁMETRO

El objetivo de estimar o calcular un parámetro se refiere a obtener, a partir de los datos muestrales, un valor que represente de la mejor manera posible una característica numérica de la población de interés. Según lo señalado en Paliz Sánchez (52), Montanero Fernández (51) y López Himely (59), un parámetro es un valor fijo, aunque generalmente desconocido, que describe algún aspecto de la población, como su media, proporción, varianza o desviación estándar.

El proceso de estimación implica trabajar con una muestra representativa de la población, aplicar procedimientos estadísticos adecuados y, a partir de ellos, obtener un estimador, que es la expresión matemática usada para aproximar el valor del parámetro. Esta estimación es de dos tipos: puntual y por intervalos.

La estimación puntual consiste en proporcionar un único valor calculado a partir de los datos muestrales como mejor aproximación al parámetro. Por ejemplo, la media muestral \bar{x} se utiliza como estimador puntual de la media poblacional μ . Este tipo de estimación, aunque directa, no refleja la incertidumbre inherente al hecho de trabajar con una muestra y no con toda la población.

La estimación por intervalos, en cambio, busca cuantificar esa incertidumbre, proporcionando un rango de valores, llamado intervalo de confianza, dentro del cual se espera que se encuentre el verdadero valor del parámetro con una determinada probabilidad o nivel de confianza (por ejemplo, 95 %). López Himely (59) resalta que este método no solo aproxima el parámetro, sino que además permite evaluar la precisión de la estimación; un intervalo estrecho indica mayor precisión, mientras que uno amplio refleja mayor incertidumbre.

Con este objetivo el investigador busca transformar la información contenida en los datos muestrales en un valor que describa a la población, considerando no solo el cálculo central, sino también la cuantificación de la incertidumbre asociada.

Influencia del tamaño de la muestral en la precisión de la estimación.

En los estudios descriptivos, el tamaño de la muestra es un elemento determinante para la calidad de las estimaciones obtenidas. Este concepto hace referencia al número de unidades de análisis incluidas en el estudio, ya sean personas, registros, objetos o mediciones.

Cuanto mayor sea el tamaño muestral, menor será la influencia del azar sobre los resultados y más estables serán los estadísticos que se calculen, como la media, la mediana, la proporción o la tasa. Por el contrario, muestras pequeñas incrementan la variabilidad aleatoria y dificultan que los valores estimados representen con fidelidad a la población.

La precisión se entiende como el grado de cercanía entre el valor estimado a partir de la muestra y el verdadero valor del parámetro en la población. Operativamente, la precisión se refleja en dos indicadores clave: el error estándar y la amplitud del intervalo de confianza.

Un error estándar reducido implica que, si se repitiera el muestreo en las mismas condiciones, las estimaciones variarían poco. Asimismo, intervalos de confianza más estrechos indican que existe menos incertidumbre alrededor de la estimación, lo que se traduce en una mayor confianza en que el valor calculado se aproxime al real.

La relación entre el tamaño de la muestra y la precisión es inversa: a medida que el número de observaciones incluidas en un estudio descriptivo aumenta, el error estándar disminuye y la estimación del parámetro poblacional se vuelve más precisa. Esto ocurre porque un mayor tamaño muestral reduce la influencia del azar y suaviza el impacto de valores atípicos o fluctuaciones extremas.

Estimación de la media de una población tiene como finalidad aproximar, a partir de una muestra, el valor central real de una variable cuantitativa continua en toda la población. Tal como señalan Montanero Fernández (2018) y Paliz Sánchez (2024), este procedimiento combina dos elementos clave: una estimación puntual y una medida de precisión que permita valorar la confianza que se puede tener en ese valor.

La estimación puntual corresponde a la media muestral, calculada como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Este valor se considera el mejor estimador de la media poblacional μ siempre que la muestra sea aleatoria y representativa. No obstante, dado que se obtiene a partir de una fracción de la población, está sujeta a variabilidad aleatoria o error de muestreo.

Para reflejar esa incertidumbre y no limitarse a un único valor, se utiliza la estimación por intervalos, que proporciona un rango de valores dentro del cual se espera que se encuentre μ con un determinado nivel de confianza, generalmente el 95 %. Este intervalo de confianza se calcula considerando la variabilidad de los datos (desviación estándar muestral s) y el tamaño de la muestra (n). Cuando la desviación estándar de la población es desconocida, que es lo habitual, se recurre a la distribución t de Student con $n-1$ grados de libertad:

$$\bar{x} \pm t_{(n-1; \alpha/2)} \cdot \frac{s}{\sqrt{n}}$$

Previo al cálculo del intervalo, es fundamental verificar dos supuestos: independencia de las observaciones y normalidad de la variable, sobre todo en muestras pequeñas. Para este fin se recomiendan métodos gráficos (histograma, cuantiles) y pruebas estadísticas como Shapiro–Wilk, más potente en tamaños reducidos, o Kolmogorov–Smirnov con corrección de Lilliefors. En muestras grandes, el teorema central del límite respalda la validez del procedimiento aun cuando la variable no sea estrictamente normal.

Estimación de la prevalencia de una población tiene como propósito determinar, a partir de datos obtenidos en una muestra, la proporción de individuos que presentan una determinada característica o condición en un momento o periodo específico. Tal como exponen González Landrián (60) y López Himely (59), esta medida es fundamental en los estudios epidemiológicos y de salud pública, ya que permite dimensionar la magnitud de un problema, identificar grupos de riesgo y orientar la planificación, evaluación e implementación de políticas o intervenciones sanitarias.

La estimación puntual de la prevalencia se obtiene calculando la proporción muestral (\hat{p}), que resulta de dividir el número de individuos con la característica de interés (x) entre el tamaño total de la muestra (n):

$$\hat{p} = \frac{x}{n}$$

Este valor es la mejor aproximación disponible del verdadero parámetro poblacional, siempre que la muestra haya sido seleccionada aleatoriamente y sea representativa de la población. La claridad en la definición de la característica estudiada y en el periodo de referencia es esencial para asegurar la validez y comparabilidad de la estimación.

Dado que toda estimación muestral está sujeta a variabilidad, se recomienda acompañar la proporción puntual con una estimación por intervalos, la cual permite establecer un rango de valores dentro del cual se espera que se encuentre la prevalencia poblacional con un determinado nivel de confianza, habitualmente del 95 %. El cálculo del intervalo de confianza más utilizado en su forma clásica se basa en la aproximación normal:

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

En este procedimiento, $Z_{\alpha/2}$ representa el valor crítico correspondiente al nivel de confianza seleccionado y el término bajo la raíz corresponde al error estándar de la proporción muestral. El intervalo resultante permite no solo presentar un valor central, sino también cuantificar la precisión de la estimación: intervalos más estrechos indican mayor precisión, mientras que intervalos amplios reflejan mayor incertidumbre.

2.6. OBJETIVO CONTRASTAR O COMPARAR CON UN PARÁMETRO

Se enmarca en la estadística inferencial y consiste en evaluar, mediante procedimientos formales, si el valor observado en una muestra es compatible o no con un valor específico del parámetro poblacional establecido como referencia. De acuerdo con Montanero Fernández (51), Molina Arias (61,62) y Ortega Páez (63,64), este proceso implica formular y someter a prueba una hipótesis estadística, que se expresa en términos del parámetro de interés, como la media (μ), la proporción (\hat{p}) o la varianza (σ^2).

El procedimiento parte del establecimiento de dos hipótesis contrapuestas: la hipótesis nula (H_0), que afirma que el parámetro poblacional es igual al valor de referencia, y la hipótesis alternativa (H_1), que sostiene que dicho parámetro es distinto, mayor o menor que ese valor, dependiendo del planteamiento del problema. Posteriormente, se selecciona un estadístico de prueba apropiado al tipo de dato y de parámetro, se calcula a partir de los datos muestrales y se compara con la distribución teórica correspondiente (normal, t de Student, ji-cuadrado, etc.) para obtener un valor de p.

En la práctica, si el objetivo es contrastar una media poblacional con un valor conocido, se emplea la prueba t de Student para una muestra cuando la desviación estándar poblacional es desconocida. En el caso de comparar una proporción con un valor teórico, se utiliza la prueba z para proporciones o, con muestras pequeñas, métodos exactos como el test binomial. Para varianzas, la comparación se realiza habitualmente con la prueba ji-cuadrado (χ^2).

El **valor de p** obtenido indica la probabilidad de observar, bajo el supuesto de que H_0 es verdadera, una diferencia igual o mayor que la encontrada en la muestra. Si esta probabilidad es menor que el nivel de significación predefinido (habitualmente $\alpha=0,05$), se rechaza la hipótesis nula y se concluye que existen evidencias suficientes para sostener la hipótesis alternativa.

2.7. EL RITUAL DE LA SIGNIFICANCIA ESTADÍSTICA

Es un procedimiento secuencial de cinco pasos que guía el contraste de hipótesis en la investigación científica. Su desarrollo integra componentes heurísticos (el planteamiento de hipótesis, la fijación del nivel de significancia y la selección de la prueba estadística) que dependen del juicio, la creatividad y las convenciones disciplinares del investigador, con un componente normativo (el cálculo del estadístico y del p-valor) que se ejecuta mediante reglas matemáticas reproducibles; y culmina en una decisión inferencial, que combina la regla formal de comparación (p frente a α).

1. Planteamiento de hipótesis

Hipótesis nula (H_0): no hay diferencias o relaciones; las cosas ocurren como siempre o se mantienen iguales.

Hipótesis alterna (H_1): sí hay diferencias o relaciones que vale la pena comprobar con los datos.

2. Fijación del nivel de significancia (α)

Representa la probabilidad máxima aceptada de rechazar H_0 cuando en realidad es verdadera (error tipo I). Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3. Selección de la prueba estadística

Según: nivel investigativo, objetivo estadístico, tipo de variable, tipo de estudio, diseño del estudio y comportamiento de los datos: χ^2 (ji-cuadrado), t de Student, ANOVA, etc.

4. Cálculo del estadístico y del p-valor

El p-valor indica la probabilidad de encontrar un resultado igual o más extremo que el observado, en el supuesto de que la hipótesis nula (H_0) sea cierta.

5. Decisión inferencial

$p < \alpha \rightarrow$ evidencia suficiente para rechazar H_0 .

$p \geq \alpha \rightarrow$ no se rechaza H_0 ; no hay evidencia estadística suficiente para sostener H_1 .

Comparación con la media poblacional

1. Planteamiento de hipótesis

Hipótesis nula (H_0): la media poblacional (μ) es igual al valor de referencia (μ_0).

Hipótesis alterna (H_1): la media poblacional difiere de μ_0 (contraste bilateral)

2. Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3. Selección de la prueba estadística

Cuando la desviación estándar poblacional (σ) es desconocida, se aplica la t de Student para una muestra, calculando el estadístico:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Donde:

- \bar{x} = media muestral
- μ_0 = media poblacional de referencia
- s = desviación estándar muestral
- n = tamaño de la muestra

4. Cálculo del p-valor

Se determina la probabilidad de obtener un valor del estadístico t igual o más extremo que el observado si H_0 fuera verdadera. Este valor depende del tipo de contraste planteado (bilateral o unilateral) y se obtiene a partir de la distribución t.

5. Decisión inferencial

Si $p < \alpha$, se rechaza H_0 y se concluye que existen evidencias estadísticas para afirmar que la media poblacional difiere del valor de referencia.

Si $p \geq \alpha$, no se rechaza H_0 interpretándose que no hay evidencias suficientes para afirmar que exista una diferencia estadísticamente significativa.

Comparación con la prevalencia de la población

1. Planteamiento de hipótesis

Hipótesis nula (H_0): la prevalencia poblacional p es igual a la prevalencia teórica (p_0).

Hipótesis alterna (H_1): la prevalencia poblacional p es distinta de p_0 (contraste bilateral)

La variable de estudio es dicotómica, con categorías “sí” y “no”.

2. Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3. Selección de la prueba estadística

Se aplica la prueba de χ^2 (ji-cuadrado) de bondad de ajuste para una muestra con variable dicotómica.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Donde:

- χ^2 = estadístico ji-cuadrado.
- O = número de casos observados con la característica de interés en la muestra.
- E = número de casos esperados según la prevalencia poblacional de referencia

4. Cálculo del p-valor

Se determina a partir del valor del estadístico χ^2 y la distribución ji-cuadrado con un grado de libertad. El p-valor indica la probabilidad de encontrar una discrepancia igual o mayor que la observada si H_0 fuera cierta.

5. Decisión inferencial

Si $p\text{-valor} < \alpha$, se rechaza H_0 : hay evidencia estadística de que p difiere de p_0 .

Si $p\text{-valor} \geq \alpha$, no se rechaza H_0 : no hay evidencia estadística suficiente para afirmar que exista una diferencia significativa.

Cuadro 9: Análisis de datos para estudios de nivel descriptivo

Objetivo	Variable	Procedimientos estadísticos
Describir o caracterizar	VARIABLES NUMÉRICAS	- Medidas de tendencia central (media, mediana, moda). Medidas de dispersión (desviación estándar, varianza, error típico) - Medidas de posición (cuartiles, deciles, percentiles). Medidas de forma (asimetría y curtosis). Gráficos: histogramas, diagramas de caja y bigotes, diagramas de tallo y hojas.
	VARIABLES CATEGÓRICAS	- Frecuencias (absolutas, relativas y acumuladas). Medidas de resumen (razones, proporciones, tasas). - Gráficos: diagramas de barras, gráficos de sectores (pastel), gráficos de Pareto
Estimar o calcular un parámetro	VARIABLES NUMÉRICAS	- Estimación puntual de la media poblacional a partir de la muestra. - Intervalo de confianza para la media (habitualmente al 95 %), usando el error estándar
	VARIABLES CATEGÓRICAS	- Estimación puntual de proporción (prevalencia o porcentaje). - Intervalo de confianza para la proporción.
Contrastar o comparar con un parámetro	VARIABLES NUMÉRICAS	Comparación de la media muestral vs. media poblacional hipotética. - Prueba t de Student para una muestra (σ desconocida).
	VARIABLES CATEGÓRICAS	Comparación de la proporción observada vs. proporción teórica esperada. - Prueba χ^2 (ji-cuadrado) de bondad de ajuste (o test binomial exacto en muestras pequeñas).

NIVEL
RELACIONAL
ANÁLISIS ESTADÍSTICO BIVARIADO

En este libro, el nivel relacional se entiende en su concepción clásica como un análisis bivariado sustentado en una matriz de datos recolectada mediante protocolo. Se trabaja con una variable de supervisión (central) y variables asociativas (vinculantes). Este nivel busca constatar vínculos estadísticos sin atribuir causalidad.

Conviene precisar que, en el ámbito de la ciencia de datos, es posible ampliar el nivel relacional hacia el análisis multivariado no supervisado, siempre que se cuente con datasets preexistentes. No obstante, en nuestra propuesta no se consideran modelos de “caja negra”: la ampliación solo se acepta si se conserva interpretabilidad.

La diferencia con el nivel descriptivo, es que este observa cada variable por separado, mientras que el relacional examina dos variables a la vez para identificar asociaciones. La diferencia con el nivel explicativo, es que este busca demostrar causalidad mediante análisis multivariado y control de confusores, mientras que el relacional solo constata asociaciones sin atribuirles carácter causal.

3.1. RELACIÓN ENTRE VARIABLES

Se entiende como el análisis conjunto de una variable de supervisión (central) con una o más variables asociativas, con el propósito de determinar si existe una asociación estadísticamente significativa. Este análisis permite identificar la dirección e intensidad de los vínculos, aportando evidencia de dependencia probabilística.

A diferencia del nivel explicativo, este nivel no busca establecer causalidad ni construir modelos multivariados. Su alcance se limita a constatar asociaciones mediante pruebas estadísticas bivariadas, que cuantifican la fuerza del vínculo entre las variables analizadas dentro de la matriz de datos.

La variable de supervisión (variable central) es la variable de estudio en los análisis del nivel relacional. Surge de la línea de investigación y constituye el eje sobre el cual se examina la asociación con otras variables asociativas. Puede ser categórica (ej. presencia o ausencia de enfermedad) o numérica (ej. nivel de glucosa).

Es importante aclarar que la variable de supervisión no equivale a la variable dependiente del nivel explicativo. La primera se usa en análisis bivariados, limitados a mostrar asociaciones; la segunda se emplea en modelos multivariados para establecer causalidad. Esta diferencia asegura coherencia metodológica entre niveles.

Las variables asociativas (vinculantes) son aquellas que se contrastan con la variable central. Su función es establecer vínculos estadísticos que permitan identificar asociaciones significativas dentro de la matriz de datos. Pueden ser cualitativas o cuantitativas, según el diseño y los objetivos de la investigación planteada.

A diferencia de las variables independientes propias del nivel explicativo, las asociativas no buscan demostrar causalidad ni construir modelos multivariados. Su papel se limita a constatar la existencia de relaciones estadísticas, proporcionando evidencia de asociación. Esta función las convierte en auxiliares esenciales del análisis relacional.

3.2. LA VALIDEZ EXTERNA

La validez externa se define como el grado en que los resultados obtenidos pueden generalizarse más allá de la muestra analizada, no solo a la población de referencia definida en el diseño sino a otros contextos, escenarios y momentos distintos al de la investigación original (51). Su esencia radica en garantizar que las conclusiones no se limiten al grupo particular estudiado, sino que puedan ser relevantes y útiles en entornos reales, sirviendo como base para la formulación de políticas, la planificación de intervenciones y la toma de decisiones basadas en evidencia.

Los estudios de nivel relacional buscan validez externa. Este tipo de investigaciones tienen como propósito identificar y cuantificar asociaciones entre variables, ya sea para describir patrones poblacionales, generar hipótesis u orientar acciones concretas (53). La utilidad real de estas asociaciones depende de que puedan proyectarse desde la muestra analizada hacia la población objetivo o incluso hacia otros contextos con características similares.

Si, por ejemplo, un estudio relacional establece que el tabaquismo se asocia significativamente con la aparición de enfermedad pulmonar crónica, este resultado solo tendrá valor estratégico si es representativo de la población y puede sostenerse en diferentes escenarios. De lo contrario, la relación identificada quedará limitada a un conjunto restringido de individuos, sin capacidad de orientar acciones a mayor escala. Por ello, la validez externa en los estudios relacionales no es un atributo opcional, sino una condición indispensable para que las conclusiones sean transferibles y aplicables.

El logro de una alta validez externa exige la adopción de decisiones metodológicas rigurosas desde la concepción del estudio. Un primer paso consiste en definir con precisión la población objetivo, especificando de manera clara el grupo al que se pretende en principio extrapolar los resultados. Continuando con un método de muestreo probabilístico que garantice que todos los elementos de la población tengan la misma probabilidad de ser incluidos en la muestra.

3.3. VARIABLES FIJAS Y ALEATORIAS

Una variable fija es aquella cuyas categorías han sido definidas deliberadamente por el investigador y su distribución es conocida antes de recolectar los datos. No se pretende generalizar estos niveles a un conjunto mayor. Por su parte, una variable aleatoria es aquella cuyos valores se obtienen mediante observación, están sujetos a variabilidad natural y representan una distribución estadística en la población.

Modelo 1. Una variable fija y una variable aleatoria. El objetivo es determinar si los valores de la variable aleatoria presentan diferencias estadísticamente significativas según los niveles establecidos de la variable fija. Este enfoque es común en los estudios relacionales orientados a comparar grupos definidos frente a una medición que varía de forma natural en la población.

Modelo 2. Dos variables aleatorias. En este escenario, ambas variables se registran tal como se presentan en la muestra, sin intervención o control directo del investigador sobre sus valores. El interés analítico se centra en determinar la existencia, dirección y fuerza de una posible asociación entre ellas.

Modelo 3. Dos variables fijas. Cuando las dos variables son fijas y categóricas, sus niveles o categorías se han definido en el diseño del estudio, y el objetivo es determinar si existe asociación entre ellas. Si la tabla de contingencia es pequeña, por ejemplo, de 2×2 , la Prueba exacta de Fisher es la más apropiada, ya que no depende de aproximaciones asintóticas y proporciona un valor p exacto.

Cuadro 10: Modelos estadísticos: variables fijas y aleatorias

Variable 1 \ Variable 2	Categórica fija	Categórica aleatoria
Categórica fija	Test de Fisher	χ^2 (ji-cuadrado) de homogeneidad
Categórica aleatoria	χ^2 (ji-cuadrado) de homogeneidad	χ^2 (ji-cuadrado) de independencia

3.4. OBJETIVO COMPARAR GRUPOS INDEPENDIENTES

Una variable fija y otra aleatoria. El objetivo de comparar grupos independientes es identificar si existen diferencias estadísticamente significativas en una variable de resultado (aleatoria) entre dos o más grupos definidos por una característica de las unidades de estudio (fija) (65). En este tipo de estudios, los grupos surgen de condiciones naturales o de características intrínsecas de los participantes.

La comparación se centra en determinar si la distribución de la variable aleatoria difiere significativamente entre los niveles de la variable fija.

La t de Student para muestras independientes es adecuada para comparar las medias de una variable continua entre dos grupos independientes. La hipótesis nula plantea que las medias son iguales, mientras que la alternativa sugiere diferencias significativas (40,65). El estadístico se calcula considerando las medias, varianzas y tamaños de muestra de los grupos.

Mientras el χ^2 (ji-cuadrado) de homogeneidad es aplicada a los datos categóricos, la t de Student se aplica a las variables continuas. Sin embargo, cada una tiene limitaciones. La prueba de χ^2 (ji-cuadrado) puede ser sensible a tamaños pequeños de muestra y no proporciona información sobre la magnitud de las diferencias, mientras que la t de Student requiere homogeneidad de varianzas y asume que los datos siguen una distribución normal (66).

El χ^2 (ji-cuadrado) de homogeneidad se utiliza para comparar proporciones entre dos o más grupos independientes y evaluar si las distribuciones de una variable categórica son similares en diferentes grupos. En esta prueba, se plantea una hipótesis nula que sostiene que no hay diferencias en las proporciones entre los grupos, y una alternativa que indica la existencia de diferencias significativas. El estadístico se calcula comparando las frecuencias observadas y esperadas, rechazándose la hipótesis nula si el p-valor es menor al nivel de significancia establecido (40,65).

Comparación de promedios entre grupos

1. Planteamiento de hipótesis

$H_0: \mu_1 = \mu_2$ no existen diferencias entre los promedios poblacionales

$H_1: \mu_1 \neq \mu_2$ existe al menos una diferencia real entre los promedios

2. Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3. Selección de la prueba estadística (t de Student)

Dos grupos independientes \rightarrow t de Student para muestras independientes.

Más de dos grupos independientes \rightarrow ANOVA de un factor.

$$t_o = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{Sp^2}{n_1} + \frac{Sp^2}{n_2}}}$$

- \bar{X}_1, \bar{X}_2 : medias muestrales de los grupos 1 y 2.
- S_1, S_2 : desviaciones estándar muestrales de los grupos 1 y 2.
- n_1, n_2 : tamaños muestrales de los grupos 1 y 2.
- Sp : desviación estándar agrupada
- t : estadístico de prueba.

4. Cálculo del p-valor

Probabilidad de rechazar H_0 cuando en realidad es verdadera.

2. Decisión inferencial

Si $p < \alpha$: se rechaza la hipótesis nula y se concluye que existen diferencias estadísticamente significativas en las medias entre los grupos.

Si $p \geq \alpha$: no se rechaza la hipótesis nula, interpretándose que no hay evidencia estadística suficiente para afirmar que las medias difieren.

Comparación de proporciones entre grupos

1) Planteamiento de hipótesis

Hipótesis nula (H_0): las proporciones son iguales entre los grupos que se comparan.

Hipótesis alterna (H_1): las proporciones difieren entre los grupos que se comparan.

2. Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3. Prueba estadística (χ^2 de Pearson)

Calcular el valor esperado para cada celda de la tabla de contingencia.

Aplicar

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Donde:

- O = valor observado
- E = valor esperado

Grados de libertad = (número de filas - 1) \times (número de columnas - 1).

4. Cálculo del p-valor

Probabilidad de rechazar H_0 cuando en realidad es verdadera.

5. Decisión inferencial

Si $p < \alpha \rightarrow$ se concluye que las proporciones comparadas son diferentes de forma estadísticamente significativa.

Si $p \geq \alpha \rightarrow$ no hay evidencia suficiente para concluir que las proporciones comparadas sean diferentes.

3.5. OBJETIVO COMPARAR MEDIDAS REPETIDAS

Comparar medidas repetidas significa analizar si existen diferencias estadísticamente significativas en una variable de interés cuando se evalúa más de una vez en los mismos sujetos, sin que el investigador manipule necesariamente las condiciones. En un estudio observacional los cambios observados responden a la evolución natural de los participantes o a exposiciones que ocurren en su entorno.

Variable categórica fija: representa el momento de medición o la condición observada (por ejemplo, “antes de la temporada de lluvias” y “después de la temporada de lluvias”). Sus niveles son definidos por el investigador para el análisis.

Variable aleatoria: corresponde al resultado medido, que puede ser cuantitativo (por ejemplo, niveles de hemoglobina) o cualitativo dicotómico (por ejemplo, presencia/ausencia de una enfermedad).

Pruebas estadísticas según el tipo de variable aleatoria

Variable aleatoria numérica

Dos mediciones: t de Student para muestras relacionadas (paired t-test), comparando la media de las diferencias con cero.

Más de dos mediciones: ANOVA para medidas repetidas.

Si no se cumplen supuestos: prueba de Wilcoxon (dos mediciones) o prueba de Friedman (tres o más mediciones).

Variable aleatoria categórica dicotómica

Dos mediciones: prueba de McNemar, que detecta cambios en las respuestas dicotómicas entre dos observaciones en los mismos individuos.

Más de dos mediciones: prueba Q de Cochran, que evalúa diferencias en respuestas dicotómicas a lo largo de tres o más mediciones.

Comparación de promedios en medidas repetidas

1. Planteamiento de hipótesis

Hipótesis nula (H_0): la media de las diferencias entre las dos mediciones es igual a cero; no existe cambio en la variable medida entre los dos momentos u observaciones.

Hipótesis alternativa (H_1): la media de las diferencias entre las dos mediciones es distinta de cero; sí existe cambio en la variable medida.

2. Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3. Prueba estadística (t de Student)

Dos mediciones relacionadas \rightarrow t de Student para muestras apareadas.

Más de dos mediciones relacionadas \rightarrow ANOVA para medidas repetidas.

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

\bar{d} : media de las diferencias entre la primera y segunda medición.

s_d : desviación estándar de las diferencias.

n: número de pares de mediciones (número de sujetos).

t: estadístico de prueba

4. Cálculo del p-valor

Probabilidad de rechazar H_0 cuando en realidad es verdadera.

5. Decisión inferencial

Si $p < \alpha$: se rechaza la hipótesis nula y se concluye que existe una diferencia estadísticamente significativa entre las dos mediciones.

Si $p \geq \alpha$: no se rechaza la hipótesis nula; no hay evidencia estadística suficiente para afirmar que las mediciones difieran.

Comparación de proporciones en medidas repetidas

1) Planteamiento de hipótesis

Hipótesis nula (H_0): la proporción de individuos con la característica de interés es igual en todos los momentos de medición.

Hipótesis alterna (H_1): la proporción de individuos con la característica de interés difiere en al menos uno de los momentos de medición.

2) Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3) Prueba estadística (McNemar)

Dos mediciones: se utiliza la prueba de McNemar.

Más de dos mediciones: se aplica la prueba Q de Cochran, que extiende el McNemar.

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

b: número de sujetos que cambiaron de respuesta positiva a negativa.

c: número de sujetos que cambiaron de respuesta negativa a positiva.

La corrección “-1” es la corrección de continuidad de Yates para tablas 2x2 con datos apareados.

4) Cálculo del p-valor

Probabilidad de rechazar H_0 cuando en realidad es verdadera.

5) Decisión inferencial

Si $p < \alpha$: se rechaza la hipótesis nula y se concluye que existe un cambio estadísticamente significativo en las proporciones entre los momentos de medición.

Si $p \geq \alpha$: no se rechaza la hipótesis nula; no hay evidencia estadística suficiente para afirmar que las proporciones difieren.

3.6. OBJETIVO ASOCIAR Y CORRELACIONAR (Ambas variables aleatorias)

Cuando estudio trabaja con dos variables el concepto de relacionar se convierte en un eje integrador. Relacionar no se limita a un tipo específico de análisis, sino que incluye a todos los procedimientos estadísticos que involucran dos variables, ya sean de naturaleza categórica, numérica o una de cada una.

Relacionar es el término que abarca diversas formas de examinar cómo interactúan las variables entre sí. Su propósito es identificar y cuantificar el grado en que una variable está vinculada con la otra, considerando la escala de medición y el contexto de la investigación. Al hablar de “relacionar”, se integran tanto las comparaciones de grupos, como las asociaciones y correlaciones, así como otras medidas que evalúan dependencia, concordancia o fuerza de vínculo entre dos variables.

Asociar, significa examinar si la distribución de una variable categórica cambia en función de las categorías de otra variable categórica. El interés no recae en valores numéricos, sino en la coincidencia, la dependencia o el patrón de distribución entre las distintas clases o grupos que forman cada variable (66).

Correlacionar es observar si, al variar una, la otra tiende a variar de manera sistemática, ya sea en la misma dirección (relación positiva) o en dirección opuesta (relación negativa). Aquí la atención se dirige a los valores individuales de cada unidad de análisis y a la coherencia de sus variaciones conjuntas (53).

Relacionar es el concepto paraguas que agrupa todos los métodos estadísticos bivariados; asociar se reserva para el vínculo entre categorías; y correlacionar para el vínculo entre unidades numéricas. Esta distinción es útil porque permite, desde la planeación metodológica, elegir el procedimiento más adecuado para el tipo de datos disponibles.

Asociar categorías

1) Planteamiento de hipótesis

Hipótesis nula (H_0): no existe asociación entre las dos variables categóricas; las categorías son independientes.

Hipótesis alterna (H_1): existe asociación entre las dos variables categóricas; las categorías no son independientes.

2) Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3) Prueba estadística (χ^2 de independencia)

La prueba a utilizar es el χ^2 (ji-cuadrado) de independencia.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Donde:

- O = valor observado
- E = valor esperado

Grados de libertad = (número de filas - 1) \times (número de columnas - 1).

4) Cálculo del p-valor

En el χ^2 (ji-cuadrado), se obtiene comparando el estadístico calculado con la distribución χ^2 (ji-cuadrado) con $(r-1) \times (c-1)$ grados de libertad.

5) Decisión inferencial

Si $p < \alpha$: se rechaza H_0 y se concluye que existe una asociación estadísticamente significativa entre las variables categóricas.

Si $p \geq \alpha$: no se rechaza H_0 ; no hay evidencia estadística suficiente para afirmar que exista asociación.

Correlacionar unidades

1) Planteamiento de hipótesis

Hipótesis nula (H_0): $\rho=0$, no existe correlación en la población; las variables son independientes linealmente.

Hipótesis alternativa (H_1): $\rho \neq 0$, existe correlación en la población; las variables presentan una relación lineal diferente de cero.

2) Fijación del nivel de significancia (α)

Valor convencional: $\alpha=0,05$ (confianza del 95 %).

3) Prueba estadística (Correlación de Pearson)

Para relación lineal entre variables numéricas continuas con distribución normal

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}}$$

x_i, y_i : valores observados.

\bar{x}, \bar{y} : medias muestrales.

r: coeficiente muestral de Pearson.

ρ : coeficiente poblacional de Pearson.

4) Cálculo del p-valor

Se transforma r en un estadístico t:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

Con $n-2$ grados de libertad. El valor p se obtiene a partir de la distribución t de Student.

5) Decisión inferencial

Si $p < \alpha$, se rechaza H_0 : existe evidencia estadísticamente significativa de correlación lineal entre las variables.

Si $p \geq \alpha$, no se rechaza H_0 : no hay evidencia suficiente para afirmar que exista correlación lineal significativa.

3.7. OBJETIVO MEDIR LA FUERZA DE ASOCIACIÓN O CORRELACIÓN

Permiten cuantificar relaciones entre variables y comprender cómo estas interactúan en diferentes contextos poblacionales. Estas herramientas no solo ayudan a identificar patrones, sino también proporcionan evidencia para diseñar intervenciones eficaces. Entre las medidas más utilizadas se encuentran el Riesgo Relativo (RR), el Odds Ratio (OR) y el coeficiente de correlación de Pearson (R).

El Riesgo Relativo (RR) es una medida de asociación que compara la probabilidad de un evento entre dos grupos (66,67). Esta herramienta se emplea ampliamente en estudios de cohorte, donde se evalúa si una exposición aumenta o disminuye el riesgo de un resultado específico. Un RR mayor a 1 indica un mayor riesgo en el grupo expuesto, mientras que un RR menor a 1 sugiere un efecto protector de la exposición.

El Odds Ratio (OR), es otra medida de asociación, utilizada principalmente en estudios de casos y controles. Esta medida estima las probabilidades relativas de exposición entre casos (personas con la condición de interés) y controles (personas sin dicha condición). Un OR mayor a 1 indica una mayor probabilidad de exposición en los casos, mientras que un OR menor a 1 sugiere una menor probabilidad (67).

Por ejemplo, en un estudio que analiza la relación entre el consumo de tabaco y el cáncer de pulmón, el RR podría calcular el riesgo de desarrollar cáncer de pulmón en fumadores frente a no fumadores en un diseño de cohorte. De manera similar, el OR podría estimar las probabilidades de que los casos (personas con cáncer de pulmón) hayan sido fumadores en comparación con los controles (sin cáncer).

El coeficiente de correlación de Pearson (R) mide la fuerza y la dirección de la relación lineal entre dos variables continuas. Un valor de R igual a 1 indica una relación lineal perfecta positiva, mientras que un valor de R igual a -1 indica una relación lineal perfecta negativa. Un valor cercano a 0 sugiere que no existe una relación lineal entre las variables (66).

Medir la fuerza de asociación

Una vez que se ha establecido que existe una asociación estadísticamente significativa entre dos variables categóricas, generalmente a través de la prueba de χ^2 (ji-cuadrado) de independencia, el siguiente paso consiste en cuantificar esa asociación, es decir, medir su fuerza. Mientras que el p-valor obtenido en la prueba de hipótesis indica si el vínculo observado es poco probable de ocurrir por azar, no ofrece información sobre la magnitud o intensidad de dicha relación. Para ello se utilizan medidas específicas que permiten describir el grado de dependencia entre las categorías.

Coeficiente de contingencia, V de Cramer, Phi (ϕ): son medidas simétricas de asociación derivadas del valor de χ^2 (ji-cuadrado), útiles para tablas de contingencia de distinto tamaño. El V de Cramer, por ejemplo, toma valores entre 0 (ninguna asociación) y 1 (asociación perfecta), independientemente del tamaño de la tabla.

Riesgo Relativo (RR): cuantifica cuántas veces más probable es que ocurra un evento en el grupo expuesto que en el no expuesto. Se interpreta como la razón entre la probabilidad del evento en expuestos y la probabilidad en no expuestos. Un RR igual a 1 indica ausencia de asociación; mayor que 1 indica que la exposición aumenta la probabilidad del evento; menor que 1 indica que la exposición es protectora.

Odds Ratio (OR): expresa la razón de las probabilidades de ocurrencia del evento entre expuestos y no expuestos. Es ampliamente utilizado en estudios de casos y controles, donde no es posible calcular riesgos directamente. Un OR = 1 indica independencia; valores superiores a 1 señalan asociación positiva y menores a 1, asociación negativa o protectora.

Índice Kappa de Cohen: específico para evaluar concordancia entre dos observadores o instrumentos en variables categóricas, descontando el acuerdo esperado por azar. Sus valores se interpretan en rangos que van desde acuerdo insignificante (cercano a 0) hasta muy bueno (cercano a 1).

Medir la fuerza de correlación

En el análisis de dos variables numéricas continuas, una vez comprobada la existencia de una correlación estadísticamente significativa, por ejemplo, mediante la prueba de hipótesis aplicada al coeficiente de Pearson, el paso siguiente consiste en evaluar la fuerza de esa correlación. Tal como indican Espinoza Pajuelo (53) y Paliz Sánchez (52), el p-valor obtenido en la prueba señala correlación, pero no describe la magnitud del vínculo, por lo que se requiere una medida específica que cuantifique esa intensidad.

La medida más utilizada es el coeficiente de correlación de Pearson (r) cuando ambas variables siguen una distribución aproximadamente normal y la relación es lineal (51). Este coeficiente varía de -1 a 1 :

- $r = 1$ indica una correlación lineal positiva perfecta: a medida que una variable aumenta, la otra también lo hace en proporción constante.
- $r = -1$ indica una correlación lineal negativa perfecta: a medida que una variable aumenta, la otra disminuye proporcionalmente.
- $r = 0$ indica ausencia de relación lineal.

Cuando las variables no cumplen los supuestos de normalidad o la relación no es estrictamente lineal, se utilizan medidas no paramétricas como la rho de Spearman o la tau de Kendall, basadas en los rangos de los datos (68). Estas medidas permiten detectar asociaciones monótonas y ofrecen una interpretación similar: valores cercanos a ± 1 indican una relación fuerte y valores cercanos a 0 una relación débil o inexistente.

Medir la fuerza de correlación permite comparar resultados entre diferentes estudios y evaluar la relevancia práctica de la relación encontrada (54). Así, un resultado estadísticamente significativo con una correlación muy débil puede tener escaso interés práctico, mientras que una correlación moderada o fuerte puede orientar decisiones estratégicas o el desarrollo de modelos predictivos.

Cuadro 11: Procedimientos estadísticos de Nivel Relacional

Objetivo	Propósito	Variabes	Procedimiento estadístico
Comparación (Una variable fija y otra aleatoria)	De grupos independientes	Numérica con distribución normal	<u>Dos grupos</u> : t de Student para muestras independientes <u>Más de dos grupos</u> : ANOVA con un factor INTERsujetos
		Numérica sin distribución normal	<u>Dos grupos</u> : U Mann-Whitney <u>Más de dos grupos</u> : H Kruskal-Wallis
		Categoría	<u>Dos o más grupos</u> : χ^2 (ji-cuadrado) de homogeneidad
	De medidas repetidas	Numérica con distribución normal	<u>Dos medidas</u> : t de Student para muestras relacionadas <u>Más de dos medidas</u> : ANOVA con un factor INTRA s ujetos
		Numérica sin distribución normal	<u>Dos medidas</u> : Wilcoxon <u>Más de dos medidas</u> : Friedman
		Categoría dicotómica	<u>Dos medidas</u> : McNemar <u>Más de dos medidas</u> : Q de Cochran
Asociación y Correlación (Ambas variables aleatorias)	Asociación	Categorías (dicotómicas o politómicas)	χ^2 (ji-cuadrado) de independencia Análisis de correspondencias
	Correlación	Numéricas con distribución normal	r de Pearson (lineal)
		Numéricas sin distribución normal	rho de Spearman (por rangos)
		Ordinales (r x r)	tau-b de Kendall
		Ordinales (r x c)	tau-c de Kendall
Medidas de asociación y correlación	Fuerza de asociación	Categorías	Riesgo Relativo (RR) Odds Ratio (OR) Kappa de Cohen
	Fuerza de correlación	Numéricas con distribución normal	r de Pearson (magnitud de relación lineal)
		Numéricas sin distribución normal	rho de Spearman (magnitud de relación por rangos)

3.8. EL NIVEL RELACIONAL NO ES CORRELACIONAL

Precisión terminológica en estadística para la investigación científica.

El nivel relacional abarca todo estudio que analiza relaciones entre variables mediante técnicas bivariadas (comparaciones, asociaciones, concordancias y correlaciones). En cambio, lo correlacional se restringe al cálculo de coeficientes estadísticos (Pearson, Spearman) aplicados únicamente a variables continuas.

Usar “correlacional” para referirse a todo lo relacional es un error metodológico grave. Como puntualizan Osada y Salvador-Carrillo (69), correlación es un procedimiento estadístico, mientras que relación es un nivel investigativo que incluye múltiples métodos de asociación. Hablar de “estudio correlacional” cuando en realidad se aplicó un χ^2 o un OR genera confusión, esos análisis no son correlaciones en sentido estricto.

La filosofía de la ciencia respalda la necesidad de precisión terminológica. Bunge (70) advirtió que la vaguedad conceptual es enemiga de la exactitud, y que la ciencia debe redefinir o sustituir términos cuando dejan de ser adecuados. De igual modo, Kuhn (71) mostró que los cambios de paradigma exigen revisar el vocabulario científico. En este caso, relacional sustituye a correlacional como denominación del nivel.

El cambio no es semántico menor, sino un ajuste epistemológico. Relacional describe la naturaleza de los estudios no experimentales que buscan dependencia entre variables, mientras que correlación es solo una de las técnicas posibles. Reservar correlacional para el procedimiento estadístico evita la falsa impresión de que todo análisis relacional se limita a calcular coeficientes lineales.

En conclusión, el nombre correcto del nivel es relacional, pues abarca con rigor todo el espectro de análisis bivariados. Correlacional debe reservarse a la técnica específica de medir correlaciones numéricas. No es un matiz opcional, sino una cuestión de rigor científico. Como afirma Osada y Salvador-Carrillo (69), “la relación corresponde a un nivel investigativo; la correlación, a un procedimiento estadístico”.

NIVEL
EXPLICATIVO

ANÁLISIS ESTADÍSTICO MULTIVARIADO

Los estudios de nivel explicativo buscan responder al porqué de los fenómenos. Su propósito es demostrar relaciones de causalidad entre variables, mostrando que los vínculos observados no son producto del azar, la confusión o la espuriedad, sino que se sustentan en fundamentos teóricos y evidencia empírica.

La diferencia con el nivel relacional radica en que el relacional se limita a mostrar relación entre variables, sin atribuirles causalidad. En cambio, el explicativo va más allá al buscar comprobar hipótesis causales mediante el control de variables confusoras y la aplicación de análisis estadístico multivariado.

La diferencia con el nivel predictivo radica en que el explicativo se centra en demostrar la relación causal existente entre variables en el presente, mientras que el predictivo toma ese conocimiento ya establecido para construir modelos que anticipen resultados futuros. El explicativo responde al porqué, y el predictivo al qué pasará.

4.1. CRITERIOS DE CAUSALIDAD

El nivel explicativo, propio de la ciencia empírico-analítica y de carácter protocolizado, busca establecer relaciones causales mediante observaciones rigurosas y experimentos controlados. En 1965, Austin Bradford Hill (72) propuso nueve criterios para orientar la inferencia causal: fuerza de asociación, consistencia, especificidad, temporalidad, gradiente biológico, plausibilidad biológica, coherencia, evidencia experimental y analogía. Estos criterios mantienen plena vigencia conceptual, pero su aplicación contemporánea exige alinearlos con las herramientas actuales de análisis de datos, de modo que puedan evaluarse con mayor rigor, precisión y coherencia.

Cuadro 12: Criterios de causalidad

Criterio de causalidad de Bradford Hill	Paso metodológico	Objetivo
1. Fuerza de la asociación	1. Relación estadística (<i>significancia estadística</i>)	EVIDENCIAR (<i>observacional</i>)
5. Gradiente biológico	2. Intensidad de la relación (<i>incluye dosis-respuesta</i>)	
4. Temporalidad	3. Relación temporal (<i>causa antes que efecto</i>)	
7. Coherencia	4. Hipótesis racional (<i>sustentada por la teoría</i>)	DEMOSTRAR (<i>experimental</i>)
8. Evidencia experimental	5. Manipulación (<i>de la variable independiente</i>)	
3. Especificidad	6. Aleatorización (<i>equilibrio de grupos</i>)	
2. Consistencia	7. Consistencia interna (<i>repetibilidad y reproducibilidad</i>)	PROBAR (<i>consistencia</i>)
9. Analogía	8. Consistencia externa (<i>confirmación en diferentes contextos</i>)	
6. Plausibilidad biológica	9. Identificación del mecanismo causal (<i>mecanismo de acción o daño</i>)	

4.2. ANÁLISIS MULTIVARIADO Y CAUSALIDAD

En el nivel explicativo, el objetivo central no es únicamente establecer si dos variables están asociadas, sino determinar si existe una relación causal entre ellas. Este propósito exige no solo la observación de una dependencia probabilística, sino también la eliminación de explicaciones alternativas como el azar o la asociación espuria (73).

El análisis multivariado es el conjunto de procedimientos estadísticos que permiten evaluar simultáneamente el efecto de múltiples variables independientes, sobre una variable dependiente. Su importancia en la demostración de causalidad radica en su capacidad para controlar el efecto de variables intervinientes, lo que incrementa la validez interna de las conclusiones (73,74).

En los estudios observacionales, donde la manipulación de la variable independiente no es posible, la inferencia causal depende de la capacidad de aislar el efecto neto de la exposición sobre el desenlace. Técnicas como la regresión lineal múltiple o la regresión logística, permiten ajustar los resultados por variables de confusión previamente identificadas y reduciendo la probabilidad de sesgo.

En los estudios experimentales, aunque la aleatorización y el control de grupos equilibran las covariables conocidas y desconocidas, el análisis multivariado permite verificar el balance de las variables, ajustar por pequeñas diferencias residuales y explorar interacciones entre factores. Además, en los diseños factoriales y modelos lineales generales, el análisis multivariado facilita la estimación de efectos principales e interacciones, optimizando la interpretación de los resultados (73,75).

El análisis multivariado es más que un procedimiento estadístico; es una estrategia metodológica que, integrada a un diseño riguroso, posibilita pasar de la simple asociación a la inferencia causal. Constituye, por tanto, un puente operativo entre los datos y la demostración de causalidad, alineado con los principios y criterios que sustentan el nivel explicativo de la investigación científica.

4.3. VARIABLE INDEPENDIENTE Y DEPENDIENTE

El estudio explicativo supone que las variaciones en la variable dependiente deben ser atribuibles, de manera directa, cuantificable y estadísticamente significativa, al efecto de una o más variables independientes. Estas relaciones se plantean como hipótesis causales fundamentadas en un marco teórico sólido y en evidencia empírica previa, y se validan mediante control estadístico o experimental, aplicando análisis multivariado que permita aislar el efecto neto de cada variable independiente y ajustar por factores de confusión que podrían distorsionar los resultados (76,77).

Las variables independientes pueden ser:

Manipuladas, cuando el investigador controla y modifica deliberadamente su valor o condición para evaluar el efecto sobre la variable dependiente. Este escenario es característico de los estudios experimentales, como el diseño completamente al azar o el diseño de bloques completamente aleatorizado, en los que la manipulación, la aleatorización y el control sistemático contribuyen a fortalecer la validez interna del estudio (73,75).

No manipuladas, cuando se observan tal como ocurren en la realidad, sin intervención directa del investigador. Son propias de los estudios observacionales explicativos, en los que el control se logra a través de procedimientos estadísticos como el análisis estratificado de Mantel–Haenszel o la regresión logística, con el fin de reducir sesgos y controlar variables intervinientes (78,79).

La variable dependiente representa el desenlace o fenómeno que se pretende explicar. Puede ser numérica (por ejemplo, niveles de glucosa, presión arterial) o categórica (por ejemplo, presencia/ausencia de una enfermedad, clasificación de severidad). Su definición operativa debe ser clara y medible, con procedimientos estandarizados y herramientas validadas, ya que la validez interna y la precisión de los hallazgos dependen en gran medida de la calidad de su medición (74,80).

4.4. OBJETIVO EVIDENCIAR (INFLUENCIA)

En los estudios explicativos de tipo observacional, el objetivo central es aproximarse a la inferencia causal en situaciones donde no resulta ético, factible o viable manipular la variable independiente. El investigador observa la exposición y el desenlace en la realidad, aplicando protocolos de investigación rigurosos y técnicas estadísticas que buscan reducir el sesgo y la confusión para fortalecer la validez de las conclusiones.

Dentro de este marco, las variables se organizan en dos categorías principales: independientes y dependiente. Las primeras agrupan tanto la exposición principal, entendida como posible causa, como las covariables o factores de control que también influyen en el desenlace. La variable dependiente, en cambio, corresponde al resultado cuya ocurrencia o variación se intenta explicar mediante el modelo analítico.

La variable independiente que representa la exposición o factor de interés que se plantea como posible causa del desenlace. Su análisis busca estimar si ejerce influencia directa sobre la variable dependiente, constituyendo el eje central de la hipótesis causal. Es, por lo tanto, el punto de partida para la construcción del modelo explicativo.

Las variables independientes intervinientes de control o covariables son aquellos factores adicionales que también ejercen efecto sobre el desenlace y que, de no ser controlados, pueden distorsionar la relación estudiada. Su inclusión en el análisis estadístico permite ajustar la estimación, aislar el efecto neto de la exposición principal y reducir la posibilidad de obtener asociaciones espurias o confusas.

La variable dependiente constituye el desenlace o resultado de interés, que refleja la influencia combinada de la exposición principal y las covariables. Es la expresión empírica del efecto que se pretende explicar y medir dentro del modelo. Su análisis, mediante técnicas multivariadas, permite estimar la magnitud y dirección de las relaciones causales, contribuyendo a una inferencia más sólida y coherente.

4.5. PROCEDIMIENTOS ESTADÍSTICOS MULTIVARIADOS

El análisis estadístico multivariado es una estrategia central para aproximarse a la causalidad. Su propósito es estimar la magnitud de la asociación entre una exposición y un desenlace, controlando el efecto de variables intervinientes que puedan distorsionar la relación. Entre las técnicas más empleadas para variables dependientes categóricas se encuentran el análisis de la varianza, el análisis estratificado de Mantel–Haenszel y la regresión logística binaria.

Análisis de la Varianza. Es un procedimiento que descompone la variabilidad total de la variable de interés en variabilidad entre grupos, que refleja diferencias en las medias y sugiere la posible influencia del factor de agrupación, y variabilidad dentro de los grupos, atribuible al azar o a factores no controlados. Calcula el estadístico F como el cociente entre ambas, de modo que un valor elevado indica que las diferencias entre grupos superan las esperadas por azar.

Análisis estratificado de Mantel–Haenszel. Es una técnica multivariada en sentido amplio, pues ajusta la medida de asociación por una o pocas variables de confusión categóricas, calculando un estimador común a través de estratos. Es útil en estudios de cohorte y casos y controles cuando el número de covariables es reducido, y su variable dependiente categórica. Además, permite evaluar la homogeneidad del efecto entre estratos, contribuyendo al análisis de interacción y modificadores de efecto.

Regresión logística binaria. Es el procedimiento multivariado más frecuente en estudios observacionales con desenlace dicotómico, como la presencia o ausencia de una enfermedad. Analiza la probabilidad de ocurrencia del evento en función de una o más variables independientes, expresando la asociación en términos de odds ratio ajustadas. Su fortaleza radica en que permite incluir simultáneamente variables continuas y categóricas, evaluar interacciones y estimar el efecto neto de la exposición, contribuyendo a cumplir el criterio de fuerza de la asociación de Bradford Hill.

4.6. ANÁLISIS DE LA VARIANZA (ANOVA)

Planteamiento de hipótesis

Hipótesis nula (H_0): las medias poblacionales de todos los grupos son iguales.

Hipótesis alterna (H_1): al menos una media poblacional difiere de las demás.

Tabla de ANOVA

Fuente de variación	Suma de cuadrados (SC)	Grados de libertad (gl)	Cuadrado medio (CM)	Estadístico F	p-valor
Entre grupos	SC_{Entre}	$k - 1$	CM_{Entre}	$F = \frac{CM_{Entre}}{CM_{Dentro}}$	p-valor
Dentro de grupos (Error)	SC_{Dentro}	$N - k$	CM_{Dentro}	—	—
Total	SC_{Total}	$N - 1$	—	—	—

Donde:

SC_{Entre} = Variabilidad debida al factor o variable categórica que define los grupos.

SC_{Dentro} = Variabilidad debida al azar o a factores no controlados dentro de cada grupo.

SC_{Total} = Variabilidad total observada respecto a la media general.

k = número de grupos.

N = número total de observaciones.

CM_{Entre} = cuadrado medio entre grupos (variabilidad debida al factor).

CM_{Dentro} = cuadrado medio dentro de los grupos (variabilidad debida al azar).

F = Estadístico de Fisher

p-valor = Probabilidad de que las diferencias entre medias se deban al azar

Decisión inferencial

Si $p < \alpha$: se rechaza H_0 → la variabilidad entre grupos es significativa; al menos dos medias difieren y se requiere análisis post hoc.

Si $p \geq \alpha$: no se rechaza H_0 → la variabilidad entre grupos no es significativa; no hay evidencia de diferencias reales entre las medias.

4.7. ANÁLISIS ESTRATIFICADO DE MANTEL–HAENZSEL

El análisis estratificado de Mantel–Haenszel es un procedimiento estadístico incluido en el nivel explicativo, es decir, en el análisis estadístico multivariado, que permite evaluar la asociación entre dos variables categóricas controlando el efecto de una tercera variable interviniente, a través de un proceso de estratificación.

En términos operativos, el método consiste en:

- Dividir la muestra en estratos definidos por la variable de control (por ejemplo, grupos de edad, sexo, área geográfica).
- Calcular en cada estrato una medida de asociación, generalmente el odds ratio (OR) o el riesgo relativo (RR), usando tablas 2×2 .
- Combinar estas medidas de asociación estratificadas en un único estimador ajustado, ponderando cada estrato según su tamaño muestral.

El procedimiento Mantel–Haenszel ofrece así:

- 1) Un estimador ajustado de la asociación que controla estadísticamente la variable estratificadora, reduciendo sesgos por confusión.
- 2) Una prueba de hipótesis para determinar si existe una asociación global estadísticamente significativa entre las variables de interés, después de controlar el efecto de la variable de estratificación.
- 3) Una prueba de homogeneidad para evaluar si la magnitud de la asociación es similar entre los estratos (efecto constante) o si varía (posible interacción o efecto modificador).

Esta técnica se ubica en el nivel explicativo porque:

- Trabaja con variable independiente, dependiente e interviniente.
- Calcula medidas de asociación (OR, RR) dentro de cada estrato.
- Se aplica cuando el objetivo estadístico es evidenciar y existe la necesidad de controlar una tercera variable que podría distorsionar la relación observada.

4.8. REGRESIÓN LOGÍSTICA BINARIA

La regresión logística binaria es un procedimiento estadístico de análisis multivariado que se emplea en el nivel explicativo, cuando el objetivo es evidenciar o demostrar la relación causal entre una o más variables independientes y una variable dependiente dicotómica, controlando simultáneamente el efecto de otras variables intervinientes.

Evidenciar la probabilidad de ocurrencia de un evento que tiene dos categorías mutuamente excluyentes (por ejemplo: sí / no, caso / control, éxito / fracaso) en función de una o varias variables independientes que pueden ser categóricas o numéricas.

Fundamento estadístico

En la regresión logística binaria, la variable dependiente Y es binaria (1 = presencia del evento; 0 = ausencia).

El modelo estima la probabilidad p de que Y = 1, usando la función logística.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

Donde:

β_0 : intercepto o constante del modelo.

β_i : coeficientes que representan el cambio en el logit de la probabilidad por unidad de cambio en X_i

X_i : variables independientes.

Para facilitar la interpretación, los coeficientes se transforman mediante la exponenciación:

$$OR = e^{\beta_i}$$

Así,

$OR > 1$ indica aumento en la probabilidad del evento y

$OR < 1$ indica disminución.

4.9. OBJETIVO DEMOSTRAR (EXPERIMENTAL)

El estudio experimental representa el nivel más alto de la evidencia científica, ya que permite establecer relaciones causales con un alto grado de certeza (74). Su fortaleza radica en dos principios esenciales: la manipulación de la variable independiente y el control de las condiciones en que se desarrolla la investigación. Ambos principios se integran para garantizar que los cambios observados en la variable dependiente sean consecuencia directa de la intervención realizada por el investigador (73).

La manipulación consiste en la intervención deliberada sobre la variable independiente con el propósito de modificarla en una dirección prevista por la hipótesis de investigación. Esta acción no es aleatoria ni improvisada, sino que responde a un planteamiento previo sustentado en evidencia científica y en un razonamiento lógico.

En el diseño experimental, la manipulación permite crear las condiciones necesarias para que el efecto observado pueda atribuirse con fundamento a la causa estudiada, descartando explicaciones alternativas (73).

El control es el conjunto de procedimientos metodológicos que aseguran que las diferencias encontradas en la variable dependiente se deban exclusivamente a la manipulación de la variable independiente. Implica la creación de un grupo experimental que recibe la intervención y un grupo control que no la recibe o recibe un placebo o tratamiento estándar.

Además, el control se fortalece mediante la aleatorización de las unidades de estudio, asegurando que las características intrínsecas e inmodificables se distribuyan equitativamente en ambos grupos. Esto equilibra los posibles factores de confusión y minimiza el sesgo de selección. El control también incluye la estandarización de las mediciones, la aplicación ciega de los procedimientos y la vigilancia de la adherencia al protocolo experimental.

Diseño completamente al azar (DCA)

Cada unidad tiene la misma probabilidad de recibir cualquiera de los tratamientos, lo que garantiza la independencia entre observaciones. Es apropiado cuando las unidades son homogéneas y no se espera variabilidad fuera del efecto de los tratamientos.

Planteamiento de hipótesis

Hipótesis nula (H_0): las medias de todos los tratamientos son iguales.

Hipótesis alterna (H_1): al menos una media de tratamiento es diferente.

Tabla de ANOVA

Fuente de variación	Suma de cuadrados (SC)	Grados de libertad (gl)	Cuadrado medio (CM)	Estadístico F	p-valor
Tratamientos (Entre grupos)	SC_{Trat}	$k - 1$	CM_{Trat}	$F = \frac{CM_{Trat}}{CM_{Error}}$	p-valor
Error (Dentro de grupos)	SC_{Error}	$N - k$	CM_{Error}	—	—
Total	SC_{Total}	$N - 1$	—	—	—

Donde:

k = número de tratamientos.

N = número total de observaciones.

CM_{Trat} = cuadrado medio de tratamientos (variabilidad debida al tratamiento).

CM_{Error} = cuadrado medio del error (variabilidad debida al azar).

F = Estadístico de Fisher

p-valor = Probabilidad de que las diferencias entre tratamientos se deban al azar.

Decisión inferencial

Si $p < \alpha$: se rechaza $H_0 \rightarrow$ la variabilidad entre tratamientos es significativa y al menos dos medias difieren. Se recomienda realizar análisis post hoc (Tukey, Bonferroni).

Si $p \geq \alpha$: no se rechaza $H_0 \rightarrow$ las diferencias entre tratamientos pueden deberse al azar.

Diseño de bloques completamente aleatorizado (DBCA)

Las unidades experimentales se agrupan en bloques homogéneos según una característica conocida que puede influir en la variable de respuesta (por ejemplo, edad, lote, ubicación). Dentro de cada bloque, los tratamientos se asignan de manera aleatoria. Su propósito es controlar la variabilidad entre bloques.

Planteamiento de hipótesis (Tratamientos)

Hipótesis nula (H_0): las medias poblacionales de todos los tratamientos son iguales.

Hipótesis alterna (H_1): al menos una de las medias poblacionales difiere de las demás.

Tabla de ANOVA

Fuente de variación	SC	gl	CM	F	p-valor
Tratamientos	SC_{Trat}	$k - 1$	CM_{Trat}	$CM_{\text{Trat}}/CM_{\text{Error}}$	p_{Trat}
Bloques	SC_{Bloq}	$b - 1$	CM_{Bloq}	$CM_{\text{Bloq}}/CM_{\text{Error}}$	p_{Bloq}
Error	SC_{Error}	$(k-1)(b-1)$	CM_{Error}	—	—
Total	SC_{Total}	$N - 1$	—	—	—

Donde:

k = número de tratamientos.

b = número de bloques

N = número total de observaciones.

CM_{Trat} = cuadrado medio de tratamientos (variabilidad debida al tratamiento).

CM_{Bloq} = cuadrado medio de bloques (variabilidad debida a los bloques).

CM_{Error} = cuadrado medio del error (variabilidad debida al azar).

Decisión inferencial

Si $p < \alpha \rightarrow$ se rechaza H_0 . la variabilidad entre tratamientos es significativa y al menos dos medias difieren.

Si $p \geq \alpha$: no se rechaza $H_0 \rightarrow$ Las diferencias entre tratamientos pueden explicarse por el azar.

Diseño Cuadrado Latino (DCL)

Permite controlar simultáneamente dos fuentes de variación (dos factores de bloqueo) además del tratamiento. Los tratamientos se disponen en una matriz cuadrada de $n \times n$ celdas, donde: n es el número de tratamientos. Cada tratamiento aparece exactamente una vez en cada fila (bloque horizontal) y una vez en cada columna (bloque vertical). El número de tratamientos es igual al número de niveles de cada factor de bloqueo.

Planteamiento de hipótesis (Tratamientos)

Hipótesis nula (H_0): las medias poblacionales de todos los tratamientos son iguales.

Hipótesis alterna (H_1): al menos una media poblacional difiere de las demás.

Tabla de ANOVA

Fuente de variación	SC	gl	CM	F	p-valor
Tratamientos	SC_{Trat}	$k - 1$	CM_{Trat}	$CM_{\text{Trat}}/CM_{\text{Error}}$	p_{Trat}
Filas	SC_{Filas}	$k - 1$	CM_{Filas}	$CM_{\text{Filas}}/CM_{\text{Error}}$	p_{Filas}
Columnas	SC_{Columnas}	$k - 1$	CM_{Columnas}	$CM_{\text{Columnas}}/CM_{\text{Error}}$	p_{Columnas}
Error	SC_{Columnas}	$(k-1)(k-2)$	CM_{Error}	—	—
Total	SC_{Total}	$k^2 - 1$	—	—	—

k = número de tratamientos (= número de filas = número de columnas).

N = número total de observaciones (k^2).

$CM_{\text{Tratamientos}}$ = cuadrado medio de tratamientos.

CM_{Filas} = cuadrado medio de filas.

CM_{Columnas} = cuadrado medio de columnas.

CM_{Error} = cuadrado medio del error.

Decisión inferencial

Si $p_{\text{Trat}} < \alpha$: se rechaza H_0 ; la variabilidad entre tratamientos es significativa y al menos dos medias difieren. Se recomienda realizar análisis post hoc (Tukey, Bonferroni).

Si $p_{\text{Trat}} \geq \alpha$: no se rechaza H_0 ; las diferencias entre tratamientos pueden deberse al azar.

Diseño factorial completamente al azar (DFCA)

Es un diseño experimental que estudia dos o más factores simultáneamente, cada uno con dos o más niveles, evaluando tanto sus efectos individuales (efectos principales) como sus combinaciones (interacciones). Las unidades experimentales se asignan aleatoriamente a todas las combinaciones posibles de niveles de los factores. Este diseño es eficiente para identificar interacciones y optimizar la cantidad de información obtenida con un número reducido de experimentos.

Para cada fuente (*Factor A*, *Factor B* y *A×B*), se plantean hipótesis:

Efecto del factor A

Ho: Todas las medias de A son iguales H1: Al menos una media de A difiere

Efecto del factor B

Ho: Todas las medias de B son iguales H1: Al menos una media de B difiere

Interacción A × B

Ho: No existe interacción entre A y B H1: Existe interacción entre A y B

Tabla de ANOVA

Fuente de variación	SC	gl	CM	F	p-valor
Factor A	SC _A	a - 1	CM _A	CM _A / CM _{Error}	p _A
Factor B	SC _B	b - 1	CM _B	CM _B / CM _{Error}	p _B
Interacción A × B	SC _{AxB}	(a - 1)(b - 1)	CM _{AxB}	CM _{AxB} / CM _{Error}	p _{AxB}
Error	SC _{Error}	N-ab	CM _{Error}	—	—
Total	SC _{Total}	N-1	—	—	—

Si $p < \alpha$ para un factor: se rechaza Ho para ese factor, indicando que hay diferencias significativas entre sus niveles.

Si $p < \alpha$ para la interacción: existe interacción significativa, lo que implica que el efecto de un factor depende del nivel del otro; en este caso, se interpreta la interacción antes que los efectos principales.

Si $p \geq \alpha$: no se rechaza Ho para esa fuente; no hay evidencia suficiente de efecto significativo.

4.10. OBJETIVO PROBAR (CONSISTENCIA)

Hace referencia a la verificación de la consistencia de los hallazgos, tanto en el plano interno (repetibilidad y reproducibilidad) como en el externo (confirmación en diferentes contextos y analogía con fenómenos semejantes). Este paso se apoya en el criterio de consistencia de Bradford Hill y busca garantizar que la relación causal propuesta no sea producto de un hallazgo aislado, sino que pueda sostenerse frente a variaciones metodológicas, poblacionales y contextuales.

En términos operativos, la consistencia interna implica repetir el análisis en la misma base de datos o en subgrupos de esta, manteniendo resultados coherentes en magnitud y dirección. Esto se traduce en la obtención de estimadores estables y en la reducción de variabilidad no explicada, lo cual aumenta la confianza en la relación causal detectada.

La consistencia externa, por su parte, requiere replicar la investigación en diferentes poblaciones, escenarios o periodos de tiempo, observando nuevamente asociaciones similares. Esto confirma que la relación no depende exclusivamente de las condiciones particulares del estudio original, sino que responde a un patrón generalizable.

La analogía, aporta solidez adicional al identificar mecanismos o efectos semejantes ya demostrados en fenómenos comparables. Por ejemplo, si un fármaco de una misma familia química ha mostrado efectos adversos específicos, es plausible que otro compuesto con estructura similar genere reacciones equivalentes.

En la práctica, probar la consistencia demanda: análisis estadístico con métodos alternativos para confirmar la robustez de los resultados; replicación multicéntrica o multinacional de estudios que permitan verificar la estabilidad de las asociaciones en diferentes contextos; y comparación con la evidencia previa, a fin de establecer analogías y coherencia teórica. Este objetivo no se limita a un único procedimiento estadístico, sino que constituye una estrategia metodológica integral para afianzar la inferencia causal y aumentar la credibilidad científica del hallazgo.

Cuadro 13. Procedimientos estadísticos del nivel explicativo

Tipo de estudio	Propósito principal	Variable dependiente	Procedimiento estadístico
Observacional	Comparar medias de más de dos grupos	Numérica	Análisis de varianza (ANOVA)
Observacional	Ajustar una medida de asociación por una o pocas variables de confusión categóricas, calculando un estimador común	Categórica dicotómica	Análisis estratificado de Mantel–Haenszel
Observacional y Experimental	Estimar efecto de variables independientes sobre variable dependiente categórica	Categórica dicotómica	Análisis de regresión logística
Observacional y Experimental	Estimar efecto de variables independientes sobre variable dependiente numérica	Numérica	Análisis de regresión lineal múltiple
Experimental	Comparar medias entre grupos	Numérica	Diseño completamente al azar (DBA)
Experimental	Controlar variabilidad mediante factores de bloqueo	Numérica	Diseño de bloques completamente aleatorizado (DBCA)
Experimental	Controlar variabilidad por dos factores	Numérica	Diseño cuadrado latino (DCL)
Experimental	Evaluar efectos e interacciones de dos o más variables	Numérica	Diseños factoriales (DFA)

NIVEL PREDICTIVO

MODELAMIENTO ESTADÍSTICO PREDICTIVO

En la investigación empírico-analítica, el nivel predictivo se apoya en la causalidad previamente demostrada bajo protocolos que aseguran trazabilidad. Su finalidad es anticipar desenlaces a partir de causas verificadas. Esto es indispensable en campos como la medicina, la psicología, el derecho, la salud pública y la educación, donde las decisiones deben basarse en evidencias explicables.

En el ámbito de la ciencia de datos, el nivel predictivo trabaja con datasets ya existentes y aplica algoritmos para estimar desenlaces futuros. Aquí lo fundamental es la capacidad pronóstica del modelo, incluso sin explicación causal directa. Esta lógica es predominante en la informática, la economía, el marketing, la ingeniería y la logística.

La diferencia con el nivel explicativo es que este se centra en demostrar la relación causal, mientras que el predictivo busca proyectar resultados futuros. La diferencia con el nivel aplicativo radica en que este último evalúa procesos y resultados de una intervención concreta, mientras que el predictivo solo estima escenarios posibles.

5.1. RELACIÓN ENTRE VARIABLES

En los estudios de nivel predictivo, la relación entre variables se establece con el propósito de anticipar un resultado o estimar la probabilidad de ocurrencia de un evento a partir de un conjunto de características previamente observadas. Se parte de la premisa de que existe un vínculo ya conocido, generalmente causal, entre las variables que intervienen, lo que permite construir modelos matemáticos o estadísticos que utilicen la información disponible para generar pronósticos confiables. En este contexto, la relación no se estudia para demostrar su existencia, como en el nivel relacional o explicativo, sino para aprovecharla como base operativa en la predicción.

Se denominan **variables exógenas** o predictoras a aquellas que aportan información para anticipar el comportamiento de otra variable. Constituyen los insumos del modelo de predicción y pueden ser numéricas o categóricas. Su selección responde a criterios teóricos (basados en la evidencia y en estudios previos) y a criterios empíricos (capacidad de mejorar la precisión predictiva).

En un modelo predictivo, cada variable exógena contribuye, en mayor o menor medida, a estimar el valor esperado de la variable que se desea pronosticar. Ejemplos comunes incluyen edad, sexo, índice de masa corporal, antecedentes patológicos, hábitos de vida, factores socioeconómicos o indicadores ambientales, según el área de estudio.

La variable endógena o variable a predecir es aquella cuyo valor o probabilidad de ocurrencia se estima a partir de las variables exógenas. Representa el resultado final del modelo y debe estar claramente definida y medida de forma confiable.

Puede ser de naturaleza numérica (por ejemplo, presión arterial, niveles de glucosa, temperatura, ingresos económicos) o categórica (por ejemplo, presencia o ausencia de enfermedad, éxito o fracaso de un tratamiento, categorías de riesgo). La precisión de la predicción dependerá de la relevancia y calidad de las variables predictoras, así como de la robustez del modelo construido.

5.2. MODELOS DE ECUACIONES ESTRUCTURALES (SEM)

Es una metodología estadística multivariada que integra en un solo marco analítico las capacidades del análisis factorial y de la regresión múltiple, permitiendo examinar de forma simultánea un conjunto de relaciones de dependencia entre variables observadas y variables latentes. A diferencia de los modelos estadísticos tradicionales, que suelen analizar las relaciones de forma aislada, el SEM admite la estimación conjunta de múltiples ecuaciones (81,82).

En el SEM, las variables observadas corresponden a aquellas medidas de forma directa a través de instrumentos, encuestas o registros, mientras que las variables latentes representan constructos teóricos que no se miden directamente, sino que se infieren a partir de indicadores observados. El modelo distingue entre variables exógenas, aquellas que actúan como predictores y no reciben influencia de otras variables dentro del sistema, y variables endógenas, aquellas que funcionan como variable a predecir y cuya variación es explicada por otras variables del modelo (83,84).

En el contexto del nivel predictivo, el SEM resulta especialmente útil porque combina la capacidad de predicción de modelos de regresión con la flexibilidad de incorporar constructos complejos y controlar el error de medición. De este modo, permite estimar cómo un conjunto de variables predictoras, observadas o latentes, contribuye a anticipar el comportamiento de una variable a predecir (85).

La validación en el SEM es un componente esencial, donde no basta con especificar ecuaciones plausibles, sino comprobar su capacidad de anticipar datos futuros con solidez. Este nivel resalta la importancia de evaluar sistemáticamente la bondad de ajuste, asegurando que el modelo no solo sea estadísticamente consistente, sino también teóricamente coherente y parsimonioso. A diferencia del nivel explicativo, donde las ecuaciones pueden servir para demostrar causalidad, en el predictivo la validez del modelo depende de su ajuste empírico y de su capacidad de generalización; de ahí que se hable de modelamiento, pues el valor del modelo reside en su validación.

5.3. MINERÍA DE DATOS (*DATA MINING*, DM)

La Minería de Datos (*Data Mining*, DM) es un proceso analítico avanzado que tiene como objetivo principal descubrir patrones, relaciones y tendencias significativas en grandes volúmenes de datos, utilizando para ello técnicas estadísticas, algoritmos de aprendizaje automático y métodos de inteligencia artificial. Su propósito va más allá de la simple organización o resumen de información: busca generar conocimiento útil y aplicable, ya sea para describir o especialmente en el nivel predictivo, anticipar comportamientos y eventos futuros (82).

El DM se estructura como una de las etapas centrales del Proceso de Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Databases, KDD), en el que, tras una fase de selección, limpieza y transformación de los datos, se aplican algoritmos que permiten extraer modelos o patrones ocultos. Estos patrones pueden adoptar diversas formas, como reglas de asociación, árboles de decisión, redes neuronales, modelos de clasificación, segmentación (clustering) o técnicas de regresión, según el objetivo planteado y la naturaleza de las variables involucradas (85,86).

En el marco del nivel predictivo, la minería de datos se centra en construir modelos que, basándose en variables exógenas o predictoras, permitan estimar el valor o la probabilidad de una variable endógena o a predecir. Esto puede incluir, por ejemplo, pronosticar el riesgo de desarrollar una enfermedad, anticipar la demanda de un producto, estimar la probabilidad de incumplimiento de pago o prever la deserción escolar (87).

Su potencial radica en la capacidad de manejar grandes volúmenes de datos, estructurados o no estructurados, y en detectar relaciones no lineales o complejas que podrían pasar inadvertidas mediante métodos estadísticos tradicionales. Además, los modelos derivados del DM pueden actualizarse con datos nuevos, mejorando progresivamente su precisión predictiva (86).

5.4. ANÁLISIS DE SUPERVIVENCIA

El análisis de supervivencia es un conjunto de técnicas estadísticas diseñadas para estudiar el tiempo transcurrido hasta la ocurrencia de un evento de interés, incorporando adecuadamente la presencia de censura y la distribución asimétrica propia de las variables de tiempo (86). El evento no se limita a la muerte, sino que puede referirse a cualquier suceso definido operacionalmente, como la aparición de una complicación, la recidiva de una enfermedad o la recuperación clínica (88).

El estimador de Kaplan-Meier constituye la técnica no paramétrica más utilizada para estimar la función de supervivencia, considerando tanto los eventos ocurridos como los datos censurados. Calcula la probabilidad acumulada de supervivencia en cada punto temporal donde se produce un evento, multiplicando las probabilidades de supervivencia de los intervalos previos. Su representación gráfica, en forma de curva escalonada, permite visualizar el patrón temporal de la supervivencia y estimar la probabilidad de no presentar el evento en horizontes específicos (86).

La prueba de log-rank es un método no paramétrico utilizado para comparar las curvas de supervivencia entre dos o más grupos independientes. Se fundamenta en contrastar el número de eventos observados y esperados en cada grupo a lo largo del tiempo, bajo la hipótesis nula de que las funciones de supervivencia son iguales en todas las poblaciones comparadas (86). El estadístico resultante sigue una distribución χ^2 (ji-cuadrado), y su significación indica si existen diferencias globales en la supervivencia.

La regresión de Cox o modelo de riesgos proporcionales es un método semiparamétrico que evalúa el efecto de múltiples variables sobre el tiempo hasta el evento. Estima la razón de riesgos (*hazard ratio*), interpretada como la variación relativa en la tasa de ocurrencia del evento asociada a cada covariable, asumiendo que esta relación se mantiene constante en el tiempo (88). La validación del supuesto de proporcionalidad es fundamental para asegurar la validez de los resultados.

5.5. SERIES TEMPORALES

El análisis de series temporales se basa en el estudio de observaciones de una variable registradas en instantes de tiempo equidistantes, con el propósito de describir su comportamiento, identificar patrones y proyectar su evolución futura (89,90). En el contexto del nivel predictivo, estas técnicas permiten modelar la dependencia temporal de los datos para anticipar valores futuros, integrando componentes como la tendencia, la estacionalidad, el ciclo y la variación irregular.

Los modelos autorregresivos (AR) explican el valor actual de la serie como una combinación lineal de sus propios valores pasados más un término de error aleatorio. El orden del modelo (p) indica cuántos retardos se utilizan; por ejemplo, un AR(1) utiliza el valor inmediatamente anterior como predictor. La utilidad de los modelos AR radica en capturar relaciones directas entre observaciones cercanas en el tiempo, siendo especialmente eficaces cuando la autocorrelación disminuye gradualmente con el desfase (90).

Los modelos de media móvil (MA) representan el valor actual de la serie como una combinación lineal de errores aleatorios ocurridos en periodos previos. El orden del modelo (q) señala cuántos términos de error se consideran; así, un MA(1) incorpora el error del periodo anterior. Estos modelos son apropiados cuando la autocorrelación es significativa solo para pocos retardos y capturan el efecto de choques temporales de corta duración (90).

Los modelos ARIMA (Autorregresivos Integrados de Media Móvil) combinan componentes autorregresivos y de media móvil, incorporando además un proceso de integración (diferenciación) para lograr la estacionariedad de la serie. La notación ARIMA(p,d,q) indica el orden autorregresivo (p), el grado de diferenciación (d) y el orden de media móvil (q). Estos modelos son versátiles, ya que permiten representar una amplia gama de estructuras temporales y, en su extensión SARIMA, incluyen términos adicionales para modelar estacionalidades (90).

Cuadro 14: Procedimientos estadísticos del nivel predictivo

Herramientas	Descripción	Procedimientos estadísticos
Modelos de Ecuaciones Estructurales (SEM)	Relacionan variables predictoras con una o varias variables a predecir, integrando análisis factorial y modelos de regresión bajo supuestos lineales o logísticos. Permiten evaluar relaciones directas e indirectas y ajustar por variables latentes.	Modelos de regresión estructural, modelos de crecimiento longitudinal, simulación de escenarios estructurales.
Minería de Datos (DM)	Permiten detectar patrones complejos y relaciones no lineales entre múltiples predictores, adaptándose para mejorar la precisión predictiva y clasificar observaciones.	Árboles de decisión (CART), random forest, máquinas de soporte vectorial (SVM), redes neuronales profundas, gradient boosting.
Análisis de supervivencia	Predice el tiempo hasta la ocurrencia de un evento, considerando la censura de datos y ajustando por múltiples factores de riesgo.	Regresión de Cox, modelos paramétricos de supervivencia (Weibull, log-normal, log-logístico), curvas de Kaplan-Meier con validación pronóstica.
Series temporales	Proyectan valores futuros en función de datos históricos, considerando componentes como tendencia, estacionalidad y variación aleatoria.	Modelos ARIMA, SARIMA, suavizamiento exponencial de Holt-Winters, modelos bayesianos dinámicos, Prophet.

5.6. OBJETIVO PREDECIR (SEM, DM)

Predecir implica estimar la probabilidad de que ocurra un evento o el valor esperado de una variable a partir de un conjunto de características observadas, denominadas variables predictoras (82). Esta estimación puede referirse a variables categóricas, como la presencia o ausencia de una enfermedad, o a variables numéricas, como la concentración plasmática de glucosa o la presión arterial sistólica, y debe expresarse como una estimación puntual acompañada de sus intervalos de confianza. A diferencia de los promedios poblacionales, la predicción busca valores individualizados.

En el enfoque de métodos estadísticos de estimación (SEM), la predicción se realiza mediante modelos como la regresión logística para variables categóricas, la regresión lineal para variables numéricas, o modelos de conteo y modelos mixtos para datos jerárquicos o medidas repetidas (82).

Este proceso sigue etapas sistemáticas: especificación del modelo con base en criterios teóricos y empíricos, estimación de parámetros mediante técnicas como máxima verosimilitud o mínimos cuadrados, evaluación del ajuste utilizando métricas apropiadas (AUC, R^2 , etc.) y validación interna y externa para asegurar su aplicabilidad en poblaciones diferentes a la de entrenamiento.

Por su parte, el enfoque de minería de datos (DM) aplica algoritmos de aprendizaje supervisado que permiten manejar grandes volúmenes de información, detectar patrones no lineales y trabajar con variables altamente correlacionadas o heterogéneas (91). Entre estos algoritmos se incluyen los árboles de decisión, bosques aleatorios, máquinas de vectores de soporte, redes neuronales y métodos de gradient boosting.

En este caso, el énfasis metodológico recae en la preparación de los datos, la optimización de modelos mediante validación cruzada y la evaluación del rendimiento usando métricas como precisión, sensibilidad, especificidad, RMSE o log-loss, según el tipo de variable a predecir.

5.7. OBJETIVO PRONOSTICAR (ANÁLISIS DE SUPERVIVENCIA)

Pronosticar es estimar, con base en datos y modelos estadísticos, el momento probable en que ocurrirá un evento, incorporando explícitamente la dimensión tiempo en el cálculo de su probabilidad. No se limita a determinar si el suceso ocurrirá o no, sino a cuantificar cuándo sucederá, considerando las condiciones individuales y contextuales que pueden adelantar o retrasar su aparición (86,92).

Estimación no paramétrica de la función de supervivencia. Estimador de Kaplan-Meier: Calcula la probabilidad de que un individuo permanezca libre del evento a lo largo del tiempo, actualizando la estimación en cada instante en que ocurre un suceso. Método actuarial o de tablas de vida: Estima la supervivencia en intervalos de tiempo predefinidos, útil en estudios con tiempos agrupados. Permite calcular tanto la supervivencia como la función de riesgo promedio en cada intervalo (87).

Pruebas de comparación de curvas de supervivencia. Prueba de log-rank (86): Compara dos o más curvas de supervivencia otorgando el mismo peso a todos los tiempos, útil para contrastar si las diferencias en el tiempo hasta el evento son estadísticamente significativas. Prueba de Gehan (generalización de Wilcoxon): Pondera más los tiempos iniciales, útil cuando los eventos se concentran en las primeras etapas del seguimiento

Modelos semiparamétricos. Modelo de riesgos proporcionales de Cox: Permite estimar el efecto de múltiples covariables sobre el tiempo hasta el evento, sin necesidad de especificar la forma de la función de riesgo basal. Produce medidas como el *hazard ratio* y posibilita pronósticos individualizados (86).

Modelos paramétricos de supervivencia. Modelos exponenciales, de Weibull, log-normal, log-logístico, entre otros: Asumen una distribución específica para los tiempos y permiten pronosticar con precisión la mediana o el tiempo medio de supervivencia. Útiles cuando se conoce o se asume un patrón particular de riesgo (constante, creciente o decreciente).

5.8. OBJETIVO PROYECTAR (SERIES DE TIEMPO)

En el contexto del objetivo predictivo, las series de tiempo desempeñan un papel central en el proceso de proyectar, ya que constituyen el marco metodológico más robusto para extrapolar tendencias y construir escenarios futuros basados en el comportamiento histórico de una variable (93). Mientras que otros enfoques predictivos dependen en gran medida de covariables externas, las series de tiempo permiten aprovechar la propia estructura temporal de los datos para identificar patrones de evolución que se repiten o se mantienen de forma consistente a lo largo del tiempo.

Su principal contribución radica en la descomposición de la variable en componentes esenciales, lo que posibilita aislar y modelar el efecto de cada uno sobre la trayectoria futura. Esto resulta especialmente útil cuando el objetivo de la proyección es comprender no solo hacia dónde se dirige el fenómeno, sino también en qué momentos y con qué intensidad pueden presentarse fluctuaciones relevantes (93).

En este sentido, los modelos de series de tiempo, tales como ARIMA, SARIMA, modelos de suavizamiento exponencial o modelos bayesianos dinámicos, proporcionan una base cuantitativa sólida para realizar proyecciones bajo distintos horizontes temporales y con intervalos de confianza que reflejan el nivel de incertidumbre inherente (81,93). Estos modelos permiten, además, incorporar cambios estructurales en los patrones históricos, ajustando la proyección cuando se detectan alteraciones significativas en el comportamiento de la variable.

En la práctica, el uso de series de tiempo en proyectar se traduce en tres beneficios estratégicos: Ajuste de la proyección a patrones históricos reales, evitando extrapolaciones lineales que ignoran variaciones periódicas o cíclicas. Evaluación de múltiples escenarios (optimista, intermedio, pesimista) a partir de modificaciones en las condiciones iniciales o en la dinámica del sistema. Optimización de la planificación a largo plazo, al proveer estimaciones temporales precisas que facilitan la asignación progresiva de recursos.

Cuadro 15: Procedimientos estadísticos del nivel predictivo

Método / Objetivo	Predecir	Pronosticar	Proyectar
Modelos de Ecuaciones Estructurales (SEM)	Regresión estructural con variables latentes para estimar el efecto de la motivación y el apoyo familiar sobre el rendimiento escolar.	Modelo de crecimiento longitudinal para analizar cómo evoluciona la satisfacción laboral a lo largo de cinco años.	Simulación de escenarios estructurales para anticipar cambios en el rendimiento académico si se modifican factores socioeconómicos.
Minería de Datos (DM)	Árbol de decisión (CART) para clasificar pacientes según su riesgo de desarrollar diabetes.	Red neuronal recurrente (RNN) para anticipar la demanda diaria de transporte público.	Simulación Monte Carlo para evaluar la variabilidad de ingresos anuales bajo distintas condiciones de mercado.
Análisis de Supervivencia	Modelo de riesgos proporcionales de Cox para calcular el riesgo de recaída en pacientes con cáncer según características clínicas.	Estimador de Kaplan-Meier para determinar la mediana de supervivencia tras un trasplante de riñón.	Proyección de curvas de supervivencia para estimar la esperanza de vida futura de una cohorte según tendencias actuales.
Series de Tiempo	Modelo autorregresivo integrado de medias móviles (ARIMA) para predecir la temperatura del día siguiente.	Modelo estacional autorregresivo integrado de medias móviles (SARIMA) para pronosticar casos de influenza en temporada alta.	Proyección de tendencia lineal para estimar el crecimiento poblacional de una ciudad en los próximos veinte años.

5.9. INFERENCIA CAUSAL VS. MODELAMIENTO PREDICTIVO

La diferencia entre el nivel explicativo y el predictivo no es un matiz teórico, sino una diferencia observable en el análisis de datos. Ambos usan técnicas multivariadas, pero con propósitos distintos: el explicativo se centra en la inferencia causal, mientras que el predictivo se ocupa del modelamiento para anticipar escenarios futuros

Nivel explicativo. Aplica técnicas orientadas a demostrar causalidad: regresión múltiple, regresión logística, ANOVA multifactorial, Mantel–Haenszel, cohortes o casos–controles. Aquí lo esencial no es solo que las variables estén asociadas, sino probar si un factor causa o modifica un resultado. El análisis busca estimar magnitud, dirección y significancia de los efectos para responder al porqué ocurre el fenómeno.

El objeto de análisis en el nivel explicativo son las relaciones causales. Cada variable independiente se examina en función de su impacto en una dependiente, bajo control de confusores y con criterios de validez interna y externa. La meta es obtener inferencias causales válidas, es decir, demostrar mecanismos que expliquen los resultados observados.

Nivel predictivo. No se concentra en interpretar coeficientes, sino en evaluar el rendimiento de modelos. Utiliza regresiones predictivas, árboles de decisión, bosques aleatorios, redes neuronales, series temporales o máquinas de soporte vectorial. Su foco es validar el modelo con métricas como AUC, sensibilidad, especificidad o RMSE. El objetivo es claro: anticipar el qué ocurrirá.

El objeto de análisis en el nivel predictivo es el modelo predictivo en sí mismo, no la relación causal entre variables. El modelo se convierte en la entidad de estudio porque su valor está en la capacidad de generalizar y pronosticar con precisión. Mientras el explicativo explica causas, el predictivo genera pronósticos útiles y verificables, evitando la “predicción ciega” basada en correlaciones espurias.

NIVEL APLICATIVO

METODOLOGÍAS DE MEJORA CONTINUA

Los estudios de nivel aplicativo constituyen la fase en la que el conocimiento científico se traduce en acciones concretas sobre la realidad. Su objetivo central es implementar intervenciones, programas o procesos orientados a mejorar u optimizar condiciones específicas. A diferencia de explorar, describir, relacionar, explicar o predecir fenómenos, el nivel aplicativo busca generar un impacto directo, verificable y sostenible.

En este nivel, el uso de datasets provenientes de intervenciones reales resulta indispensable. Los registros sistemáticos permiten monitorear indicadores de desempeño, optimizar procesos mediante análisis estadístico y garantizar que las mejoras alcanzadas se consoliden en la práctica.

La diferencia con el nivel predictivo radica en que el predictivo tiene como objeto de análisis los modelos para anticipar desenlaces. En cambio, el aplicativo centra su objeto de análisis en la realidad intervenida, evaluando el rendimiento de las acciones implementadas y consolidando mejoras en procesos o resultados.

6.1. CONTROL ESTADÍSTICO DE PROCESOS (CEP)

Constituye un conjunto de métodos y herramientas estadísticas orientadas a la monitorización, análisis y mejora de procesos productivos o de servicios, con el objetivo de reducir la variabilidad y mantener los resultados dentro de límites de control previamente establecidos (94). Su fundamento se asienta en la distinción entre variabilidad natural o común, inherente al sistema, y variabilidad especial o asignable, atribuible a causas identificables y, por lo tanto, potencialmente eliminables (95).

En el contexto de la investigación de nivel aplicativo, el CEP adquiere un carácter instrumental para evaluar y optimizar intervenciones dirigidas a mejorar una condición específica en una población o sistema. Este nivel de investigación, cuyo fin primario es transformar la realidad mediante acciones concretas, requiere no solo documentar la implementación de las intervenciones, sino cuantificar rigurosamente su eficacia, eficiencia e impacto (96). En este marco, el CEP actúa como un sistema de retroalimentación estadística que permite:

Definir y medir variables críticas del proceso objeto de mejora, asegurando la confiabilidad de los sistemas de medición. Analizar la variabilidad para identificar causas y priorizar acciones correctivas. Implementar mejoras y documentar cambios, integrando herramientas como gráficos de control univariados o multivariados, análisis de capacidad de procesos. Establecer mecanismos de control que garanticen la sostenibilidad de los resultados alcanzados tras la intervención (97).

En proyectos de nivel aplicativo, la lógica del CEP se alinea con la estructura de metodologías de mejora continua como Seis Sigma y su ciclo DMAIC (Definir, Medir, Analizar, Mejorar y Controlar), que aportan un marco sistemático para validar estadísticamente los efectos de las acciones implementadas (95,97). Esta integración permite que los resultados no solo sean estadísticamente significativos, sino también operacionalmente relevantes para la toma de decisiones.

6.2. METODOLOGÍA SEIS SIGMA (DMAIC)

Seis Sigma es una metodología de mejora continua, cuyo propósito es reducir la variabilidad de los procesos, eliminar defectos y alcanzar un desempeño cercano a la perfección (95). Este estándar exige procesos estables, predecibles y con mínima variabilidad, lo que se logra a través del ciclo DMAIC (97).

Definir (Define). En esta fase se identifica claramente el problema, se delimitan sus alcances, se establecen los objetivos del proyecto y se determinan las necesidades y expectativas del cliente interno o externo. Se desarrollan diagramas de procesos y se documenta la situación actual, fijando indicadores clave de desempeño (95).

Medir (Measure). Consiste en recopilar datos fiables sobre las variables críticas del proceso para cuantificar su desempeño actual. Se evalúa la capacidad del sistema de medición y se calculan métricas como el DPMO (Defectos por Millón de Oportunidades) o los índices de capacidad del proceso (C_p , C_{pk}) (98).

Analizar (Analyze). Se identifican las causas raíz de la variabilidad o de los defectos mediante técnicas estadísticas (diagramas de Pareto, análisis de correlación, regresión, ANOVA) y herramientas de calidad (Ishikawa, estratificación). El objetivo es priorizar aquellas causas que tienen un mayor impacto sobre el problema (97).

Mejorar (Improve). En esta etapa se diseñan e implementan soluciones para eliminar o reducir las causas críticas detectadas. Se aplican métodos de optimización como el Diseño de Experimentos (DOE) y se validan los cambios con pruebas piloto, asegurando que las mejoras propuestas sean viables y sostenibles (95).

Controlar (Control). El propósito es mantener las mejoras logradas mediante sistemas de monitoreo continuo, gráficos de control y planes de reacción ante desviaciones. Se documentan las prácticas optimizadas y se transfiere la propiedad del proceso al responsable operativo para asegurar la estabilidad a largo plazo (98).

6.3. METODOLOGÍA CRISP-DM

La metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) es un estándar internacional para la gestión de proyectos de minería de datos y analítica avanzada, diseñado para proporcionar una guía estructurada, flexible y no propietaria que permita abordar de forma sistemática problemas complejos basados en datos (99,100).

Su origen se remonta a finales de la década de 1990, como resultado de la colaboración entre DaimlerChrysler, SPSS y NCR, y se ha consolidado como el modelo de referencia por su independencia de herramientas y su aplicabilidad en múltiples sectores.

Este enfoque se basa en la premisa de que el valor de los datos se maximiza al convertirlos en conocimiento accionable, mediante un ciclo iterativo de seis fases conectadas, en las que es posible avanzar y retroceder según las necesidades y hallazgos del proyecto (99). Características clave

- Iterativa y flexible: permite retroceder a fases anteriores para mejorar resultados.
- Independencia tecnológica: no depende de un software específico.
- Orientación a negocio: el proyecto se evalúa por su impacto en la organización, no solo por métricas técnicas.
- Enfoque en la calidad de datos: reconoce que la preparación y validación de la información son críticas para el éxito del análisis.

En el contexto de la salud pública, el CRISP-DM proporciona un marco metodológico robusto y adaptable para proyectos de analítica de datos orientados a la prevención, vigilancia y control de problemas sanitarios. Este enfoque permite integrar de forma lógica la perspectiva estratégica de gestión en salud, con las exigencias técnicas de la minería de datos necesarias para procesar grandes volúmenes de información epidemiológica, clínica y sociodemográfica.

Fases de la metodología CRISP-DM

Comprensión del negocio (Business Understanding). Se definen los objetivos y requisitos del proyecto desde la perspectiva organizacional, traduciendo estos objetivos a metas técnicas de minería de datos. Incluye la evaluación inicial de la situación, la definición del problema y la elaboración de un plan de proyecto (99).

Comprensión de los datos (Data Understanding). Se recopilan datos iniciales y se realiza una exploración preliminar para identificar su contenido, calidad y relevancia. En esta fase se detectan valores ausentes, inconsistencias, posibles sesgos y se formulan hipótesis iniciales sobre las relaciones entre variables (99).

Preparación de los datos (Data Preparation). Construir el conjunto de datos final que se utilizará en el modelado. Comprende limpieza, integración de fuentes, selección y transformación de variables, reducción de dimensionalidad y gestión de ruido, asegurando la calidad y pertinencia de la información (100).

Modelado (Modeling). Se seleccionan y aplican técnicas de modelado estadístico o de aprendizaje automático, configurando parámetros y evaluando diferentes algoritmos. Puede requerir volver a la fase de preparación para ajustar variables según las exigencias del modelo (99).

Evaluación (Evaluation). Se verifica si el modelo cumple los criterios de éxito definidos en la comprensión del negocio, analizando resultados y posibles limitaciones. En esta fase se determina si el modelo es adecuado para su implementación o si se necesitan ajustes adicionales (100).

Despliegue (Deployment). Consiste en implementar el modelo o los resultados obtenidos en el entorno de aplicación. Esto puede implicar la generación de informes, tableros de control o la integración en sistemas operativos. Se planifica el monitoreo y mantenimiento del modelo para garantizar su vigencia (99).

Cuadro 16: Metodologías de mejora continua.

Metodología	Objetivo en investigación aplicada	Procesos
Control Estadístico de Procesos (CEP)	Monitorear, analizar y mejorar procesos de intervención para reducir la variabilidad y mantener resultados dentro de límites de control, asegurando la sostenibilidad de los cambios.	<ol style="list-style-type: none"> 1. Definir y medir variables críticas. 2. Analizar la variabilidad. 3. Implementar mejoras. 4. Controlar la estabilidad del proceso.
Seis Sigma (DMAIC)	Reducir variabilidad y defectos en procesos aplicados, optimizando su desempeño y garantizando que las mejoras sean sostenibles y medibles.	<ol style="list-style-type: none"> 1. Definir: problema, alcance, objetivos, indicadores clave. 2. Medir: recopilar datos, Cp, Cpk, DPMO. 3. Analizar: causas raíz (Pareto, ANOVA, regresión). 4. Mejorar: soluciones, DOE, validación. 5. Controlar: monitoreo, gráficos de control, planes de reacción.
CRISP-DM (<i>Data Mining</i>)	Gestionar proyectos de analítica y minería de datos en salud pública y ciencias aplicadas, para convertir datos en conocimiento útil para la toma de decisiones.	<ol style="list-style-type: none"> 1. Comprensión del negocio. 2. Comprensión de los datos. 3. Preparación de datos. 4. Modelado. 5. Evaluación. 6. Despliegue.

6.4. OBJETIVO EVALUAR (INCLUYE MONITOREAR)

El objetivo Evaluar, consiste en medir de manera sistemática el estado inicial y la evolución de un proceso, intervención o sistema, con el fin de establecer una línea base y generar evidencia objetiva. Involucra también el monitoreo continuo, entendido como la observación periódica o en tiempo real de los indicadores clave para detectar variaciones, desviaciones o tendencias que requieran ajustes oportunos.

En el marco del Control Estadístico de Procesos (CEP), evaluar implica identificar la variabilidad natural y la variabilidad especial mediante herramientas como gráficos de control, análisis de capacidad de procesos y estudios descriptivos de variables críticas (94,101). El monitoreo, es esencial para garantizar la trazabilidad y la detección temprana de desviaciones respecto de los límites de control o de especificación.

En la metodología Seis Sigma – DMAIC, este objetivo se vincula principalmente con las fases Measure y Analyze. La primera asegura la recopilación de datos confiables, la validación de los sistemas de medición (MSA) y el cálculo de métricas iniciales como el nivel sigma o el DPMO (defectos por millón de oportunidades). La segunda permite analizar las causas de la variabilidad observada, utilizando herramientas como diagramas de Pareto, análisis de correlación, ANOVA o regresión múltiple (95,97).

En CRISP-DM, evaluar se asocia con las fases iniciales, donde se definen los objetivos de negocio y los criterios de éxito, se explora la estructura y calidad de los datos, y se identifican las variables relevantes para el análisis (99,100). El monitoreo, en este marco, se materializa posteriormente en la fase Deployment, mediante sistemas que permiten seguir el comportamiento de los modelos y su adecuación a los datos nuevos.

Aplicado a salud pública, este objetivo permitiría, por ejemplo, evaluar y monitorear la cobertura de una campaña de vacunación, midiendo en tiempo real los avances frente a la meta establecida, identificando zonas con menor cobertura y verificando la consistencia de los datos reportados.

6.5. OBJETIVO CALIBRAR (INCLUYE OPTIMIZAR)

El objetivo Calibrar, se orienta a ajustar la intensidad, frecuencia o configuración de una intervención con el fin de que los resultados obtenidos se sitúen dentro de los límites de tolerancia o especificación previamente establecidos. En este marco, Optimizar constituye el paso siguiente y complementario: busca maximizar su rendimiento, eficacia o eficiencia, logrando el mejor resultado posible con los recursos disponibles y reduciendo la variabilidad al mínimo técnicamente alcanzable.

En el Control Estadístico de Procesos (CEP), calibrar implica modificar parámetros operativos (por ejemplo, temperatura, velocidad, dosis, tiempo) basándose en datos provenientes de gráficos de control, estudios de capacidad del proceso (Cp, Cpk) y análisis de variabilidad. Optimizar, por su parte, se orienta a elevar la capacidad del proceso y desplazar su promedio hacia el valor objetivo, reduciendo la dispersión en torno a él (94,101).

En la metodología Seis Sigma – DMAIC, calibrar se vincula con las fases Analyze y Improve. La primera identifica la relación causa-efecto entre variables críticas de entrada y resultados, utilizando análisis estadístico multivariante, ANOVA o regresión múltiple. La segunda aplica Diseño de Experimentos (DOE), simulaciones o técnicas de optimización para encontrar la configuración óptima del proceso, verificando que las mejoras sean estadísticamente significativas y operativamente viables (95,97).

En CRISP-DM, calibrar corresponde a la fase de Data Preparation, donde se realizan transformaciones, selecciones de variables y ajustes de datos que faciliten un modelado más preciso. Optimizar se materializa en la fase de Modeling, mediante la búsqueda de la configuración de parámetros que maximice métricas de desempeño (precisión, sensibilidad, especificidad, AUC), seguida de su validación en la fase de Evaluation (99,100).

6.6. OBJETIVO CONTROLAR (INCLUYE CONSOLIDAR)

El objetivo Controlar, se orienta a mantener los resultados de un proceso o intervención dentro de límites de control y tolerancia establecidos, asegurando que las mejoras alcanzadas en fases previas se sostengan en el tiempo. En estrecha relación, Consolidar supone afianzar y fortalecer los logros alcanzados, incorporando las mejoras al funcionamiento habitual de la organización o sistema, de modo que pasen a formar parte de su estándar operativo.

En el Control Estadístico de Procesos (CEP), controlar significa aplicar de forma sistemática herramientas como gráficos de control (\bar{x} -R, \bar{x} -S, p, np, c, u), índices de capacidad a largo plazo y análisis de tendencias, para verificar que las variables críticas se mantengan en estado de “proceso bajo control estadístico” (94,101). Consolidar, desde este enfoque, supone que el proceso mantenga una variabilidad exclusivamente natural (causas comunes) y que los indicadores permanezcan estables incluso frente a cambios en la demanda o en el entorno operativo.

En la metodología Seis Sigma – DMAIC, controlar corresponde a la fase Control, donde se implementan planes de control, manuales de procedimiento, capacitación del personal y auditorías de proceso para garantizar que las mejoras no se deterioren. Consolidar se vincula con la adopción formal de las mejores prácticas, su documentación y estandarización como parte de los sistemas de gestión de calidad (95,97).

En CRISP-DM, controlar puede equipararse a la fase de Deployment, cuando se implementan soluciones en un entorno de uso real y se establecen rutinas de seguimiento del desempeño del modelo o sistema. Consolidar se logra mediante la revisión periódica de los resultados en la fase de Evaluation, confirmando que la herramienta o procedimiento sigue cumpliendo con los criterios de éxito definidos (99,100).

6.7. OBJETIVO ACEPTAR (INCLUYE VALIDAR)

El objetivo Aceptar, se orienta a determinar, mediante criterios predefinidos y evidencia estadística, si un proceso, procedimiento o intervención puede considerarse válido para su uso generalizado. Esto implica contrastar los resultados obtenidos contra límites de tolerancia o especificaciones técnicas que han sido establecidos previamente y tomar la decisión de aprobación o rechazo del sistema evaluado.

En el Control Estadístico de Procesos (CEP), aceptar implica la aplicación de planes de muestreo de aceptación, para decidir si un lote, servicio o intervención cumple los límites de especificación o tolerancia establecidos (94,101). La validación, se apoya en el cálculo de índices de capacidad (C_p , C_{pk}), análisis de conformidad y verificación de que el proceso se mantiene en control estadístico durante un periodo de prueba.

En la metodología Seis Sigma – DMAIC, aceptar se vincula con la fase Control, cuando se verifica que las soluciones implementadas han alcanzado el nivel sigma o la reducción de defectos objetivo, y se decide formalmente su adopción (95,97). Validar se articula desde la fase Measure (definición de métricas y criterios) hasta la fase Control, evaluando si la mejora se sostiene en el tiempo y en diferentes condiciones operativas.

En CRISP-DM, aceptar se relaciona con la fase Evaluation, donde se comparan los resultados del modelo o procedimiento contra los umbrales de éxito definidos en la fase de Business Understanding. Validar implica pruebas adicionales con datos independientes o en entornos piloto para asegurar que el rendimiento observado no es producto de condiciones particulares del conjunto de datos original (99,100).

Finalmente, Aceptar es la decisión final de adopción basada en la conformidad con criterios técnicos y estadísticos, mientras que Validar es el proceso analítico que confirma la idoneidad, reproducibilidad y pertinencia de la intervención o proceso evaluado. Ambos pasos son indispensables para cerrar el ciclo de la investigación aplicada y garantizar que las mejoras alcanzadas se puedan implementar con confianza.

Cuadro 17: Procedimientos estadísticos del nivel aplicativo

Objetivo	Control Estadístico de Procesos (CEP)	Seis Sigma – DMAIC	CRISP-DM
1. Evaluar y Monitorear	<ul style="list-style-type: none"> - Establecer línea base del proceso. - Identificar variabilidad natural y especial. - Uso inicial de gráficos de control. 	<p>Measure: Recopilar datos confiables.</p> <p>Analyze: Identificar causas iniciales.</p>	<p>Data Understanding: Explorar calidad y estructura de datos.</p> <p>Business Understanding: Definir métricas clave.</p>
2. Calibrar y Optimizar	<ul style="list-style-type: none"> - Ajustar parámetros del proceso para que los resultados se mantengan dentro de límites de tolerancia. - Mejorar capacidad del proceso (Cp, Cpk). 	<p>Analyze: Determinar relación causa-efecto.</p> <p>Improve: Implementar ajustes para maximizar rendimiento.</p>	<p>Data Preparation: Transformar variables para mejorar modelado.</p> <p>Modeling: Ajustar parámetros del modelo para optimizar resultados.</p>
3. Controlar y Consolidar	<ul style="list-style-type: none"> - Mantener resultados dentro de límites de control. - Monitoreo continuo de indicadores críticos. - Documentar estabilidad del proceso. 	<p>Control: Establecer planes de control, estandarizar procedimientos.</p>	<p>Deployment: Implementar sistemas de monitoreo en tiempo real.</p> <p>Evaluation: Validar estabilidad del modelo en producción.</p>
4. Aceptar y Validar	<ul style="list-style-type: none"> - Confirmar que los resultados cumplen límites de especificación o tolerancia. - Aplicar planes de muestreo de aceptación. 	<p>Measure: Comparar con criterios de aceptación.</p> <p>Control: Certificar cumplimiento antes de estandarización.</p>	<p>Evaluation: Comparar resultados contra umbrales definidos.</p> <p>Business Understanding: Revisar alineación con objetivos estratégicos.</p>

6.8. La investigación tecnológica es parte del nivel aplicativo

En la clasificación de los niveles de investigación, cada nivel se distingue por su objeto de estudio y el tipo de análisis de datos característico. Pretender añadir un “séptimo nivel tecnológico” carece de fundamento metodológico, pues no introduce un nuevo tipo de análisis, sino que se apoya en el mismo análisis de datos que define al nivel aplicativo.

Análisis en la investigación aplicada. El nivel aplicativo utiliza metodologías estadísticas de control y optimización de procesos: control estadístico de procesos (CEP), Seis Sigma, CRISP–DM y auditorías con estadística. Todas estas técnicas se centran en evaluar resultados de intervención y consolidar mejoras en la práctica real, aplicando directamente el conocimiento científico en contextos concretos.

La llamada “investigación tecnológica” no incorpora un análisis de datos distinto: también usa modelos de control, optimización, calibración y validación, que son los mismos procedimientos empleados en el nivel aplicativo. Cuando se desarrolla un producto, proceso o sistema, los datos se analizan para probar eficacia, eficiencia y capacidad de mejora continua, exactamente como ocurre en la investigación aplicada.

Objeto de análisis. Desde el punto de vista del objeto de análisis, la investigación aplicada estudia resultados de intervención en el mundo real, y la tecnológica se limita a un subconjunto de esos resultados: productos o procesos innovadores. No constituye un objeto de estudio nuevo ni un tipo de inferencia distinto, sino una variación práctica dentro del mismo marco aplicado.

Por ello, no es necesario un séptimo nivel. La investigación tecnológica no inaugura un campo metodológico diferente, sino que se integra plenamente en la investigación aplicada, donde el análisis de datos ya contempla la validación y optimización de procesos y tecnologías. Hablar de “nivel tecnológico” sería redundante y confundiría la taxonomía; lo correcto es reconocerlo como una modalidad de la investigación aplicada, no como un nivel independiente.

**PARTE III - DERIVADOS DE
LOS NIVELES DE
INVESTIGACIÓN**

1. Derivados de los niveles de investigación

El recorrido por los niveles de investigación muestra que cada uno no solo se distingue por sus objetivos y técnicas de análisis, sino que también genera elementos derivados que emergen como consecuencias metodológicas. Estos derivados aseguran la coherencia interna del proceso científico y permiten articular los diferentes componentes de la investigación.

En un sentido amplio, los principales derivados que se reconocen en cada nivel son:

- Las variables analíticas, que traducen los objetivos en dimensiones observables y medibles.
- Objetos de análisis, que definen el objeto empírico sobre el cual se focaliza el estudio.
- Hipótesis y supuestos, que orientan la comprobación de relaciones.
- Productos esperados, que sintetizan los resultados inmediatos de cada nivel.
- Criterios de calidad y validez, que permiten juzgar la solidez de los hallazgos.

No obstante, este libro ha sido concebido para mantener un enfoque preciso y práctico. Por ello, en lugar de abordar todos estos derivados, se concentrará en el más crítico y transversal: las variables analíticas.

Las variables constituyen el puente entre los objetivos de la investigación y el análisis de datos. Distinguir correctamente entre variables de estudio y variables auxiliares en cada nivel es esencial para garantizar que los objetivos se traduzcan en evidencias empíricas sólidas. En las páginas siguientes se desarrollará este aspecto con detalle, mostrando cómo cada nivel investigativo establece un patrón específico de variables predominantes.

2. Marco conceptual general de las variables analíticas

El concepto de variables analíticas se plantea como una categoría integradora que supera la dispersión terminológica en metodología de la investigación. A lo largo de la literatura, se han empleado múltiples denominaciones, independientes, dependientes, intervinientes, de control, predictoras, de estudio o de caracterización, lo que genera confusión al transitar entre niveles investigativos.

El concepto de variables analíticas busca unificar todas estas categorías bajo un mismo paraguas teórico y operativo. Se define como:

Toda dimensión observable u operativa que participa en el análisis de datos, asumiendo diferentes roles según el nivel investigativo en el que se inserte.

El carácter central de una variable radica en ser insumo del análisis, no en el nombre que reciba. Su importancia está en la función que cumple en el proceso investigativo: puede ser de caracterización en un estudio descriptivo, de supervisión en uno relacional o predictora en un modelo de pronóstico.

Desde el punto de vista epistemológico, la categoría de variables analíticas aporta tres beneficios fundamentales:

- **Carácter integrador.** Reconcilia las múltiples taxonomías de variables en una categoría común, evitando duplicidades y contradicciones terminológicas.
- **Flexibilidad metodológica.** Permite que una misma variable asuma distintos roles según los objetivos de investigación y el nivel de complejidad del análisis de datos.

- **Coherencia pedagógica.** Facilita la enseñanza y el aprendizaje de la investigación científica, al ofrecer un marco conceptual unificado que muestra cómo las variables se transforman y adquieren funciones diversas a lo largo del recorrido de una línea de investigación.

Las variables analíticas resaltan su vínculo esencial con el análisis de datos, desde lo cualitativo exploratorio hasta los modelos predictivos y aplicativos. Más que la etiqueta que reciban, lo relevante es la función que cumplen como parte del engranaje del análisis de datos, generando evidencia en coherencia con cada nivel de investigación.

Definición de Variable de estudio

La variable de estudio es la dimensión central y permanente de una línea de investigación. Representa el fenómeno, atributo o resultado que se busca observar, describir, relacionar, explicar, predecir o evaluar, según el nivel investigativo correspondiente. Su denominación cambia a lo largo de los niveles (variable de interés en el descriptivo, de supervisión en el relacional, dependiente en el explicativo, endógena en el predictivo y evaluativa en el aplicativo), pero es en esencia la misma, que se deriva de la línea de investigación: constituye el eje sobre el cual se organiza todo el análisis de datos.

Definición de Variables auxiliares

Las variables auxiliares son todas aquellas que acompañan a la variable de estudio para enriquecer el análisis, aportar contexto, contrastar asociaciones o explicar variaciones. No constituyen las variables centrales de la investigación, pero cumplen funciones específicas según el nivel: se denominan de caracterización en el descriptivo, asociativas en el relacional, independientes en el explicativo, exógenas en el predictivo y de calibración en el aplicativo. Pueden ser sociodemográficas, clínicas, ambientales, institucionales o conductuales, y su inclusión permite segmentar, ajustar, controlar o anticipar resultados.

3. Variables analíticas en el nivel exploratorio

En el nivel exploratorio, las variables analíticas se orientan al descubrimiento inicial del fenómeno de estudio. Aquí, la variable de estudio aún no se encuentra plenamente definida ni operacionalizada; está en proceso de construcción. Su función es perfilar el eje central de la investigación, clarificando su alcance y delimitando sus dimensiones fundamentales.

La variable de estudio en esta etapa comienza a consolidarse como el núcleo sobre el que girarán los niveles posteriores. No se trata de un elemento inexistente, sino de una entidad incipiente que el investigador debe reconocer, precisar y preparar para convertirse en la variable de interés en el nivel descriptivo. En términos clínicos, por ejemplo, el dolor, el malestar general, etc. vienen construyendo un diagnóstico.

En el exploratorio sí pueden existir verdaderas variables auxiliares. Estas no son simples categorías tentativas, sino atributos ya observados o medidos con claridad. Ejemplos clásicos en la práctica clínica incluyen presión arterial elevada, fiebre confirmada o colesterol en 300 mg/dl. Tales variables cumplen un papel equivalente a la observación, los testimonios o el análisis documental: actúan como insumos del descubrimiento, aportando evidencia concreta para sustentar la construcción de la variable de estudio.

Las variables auxiliares en este nivel no tienen jerarquía sobre la variable de estudio; su rol es alimentar el proceso exploratorio y ofrecer información complementaria para perfilar el problema. Funcionan como acompañantes del descubrimiento, al mismo nivel que las técnicas cualitativas de recolección de información, contribuyendo a organizar los hallazgos dispersos y a identificar posibles direcciones para la investigación.

En síntesis, el nivel exploratorio no es una instancia vacía de variables. Es el espacio donde la variable de estudio empieza a configurarse a partir de categorías emergentes y donde, en paralelo, las variables auxiliares pueden estar presentes como verdaderas dimensiones observadas o medidas, al mismo rango que la observación o el testimonio.

4. Variables analíticas en el nivel descriptivo

En el nivel descriptivo, las variables analíticas se organizan en dos categorías fundamentales: la variable de interés (variable de estudio) y las variables de caracterización (auxiliares). La primera concentra la atención en el objeto central de análisis, mientras que las segundas proporcionan un marco contextual que complementa la interpretación de los hallazgos sin desplazar el foco principal.

La variable de interés (variable de estudio), Constituye el eje del estudio descriptivo, ya que representa el fenómeno específico que se desea medir o cuantificar en la población o muestra. Es el atributo principal que responde directamente a la pregunta de investigación, funcionando como punto de referencia para todo el análisis, su definición debe ser rigurosa, coherente y operacionalizable.

Además, la variable de interés puede adoptar diferentes naturalezas. Puede ser numérica continua (niveles de glucosa, presión arterial o índice de masa corporal) o bien categórica o cualitativa (grupo sanguíneo, diagnóstico clínico o tipo de tratamiento recibido). Esta dualidad obliga al investigador a seleccionar procedimientos estadísticos apropiados que permitan describirla con precisión.

Las variables de caracterización (auxiliares), denominadas también contextuales, cumplen la función de describir el entorno en el que se observa la variable de interés. No forman parte directa del objetivo central del estudio, pero aportan información esencial para comprender la población investigada y para segmentar los análisis.

Estas variables suelen corresponder a aspectos sociodemográficos (edad, sexo, nivel educativo, ocupación), clínicos (comorbilidades, antecedentes familiares, estado nutricional), geográficos (lugar de residencia, ámbito urbano o rural) o institucionales (tipo de hospital, disponibilidad de recursos sanitarios). Aunque secundarias en términos de objetivos, son decisivas para contextualizar la variable de interés y permiten descubrir patrones o desigualdades dentro del grupo o población estudiada.

5. Variables analíticas en el nivel relacional

En el nivel relacional, las variables analíticas se dividen en dos categorías principales: la variable de supervisión (central) y las variables asociativas (vinculantes). La primera constituye el eje de observación y análisis, mientras que las segundas permiten constatar la existencia de vínculos estadísticos. La intención aquí es identificar dependencias probabilísticas entre variables, sin llegar aún a establecer causalidad.

La variable de supervisión (variable de estudio), también llamada variable central, focal o eje de análisis. Representa el núcleo de atención del investigador, quien concentra sus esfuerzos en supervisar su comportamiento en relación con otros factores. Aunque no equivale todavía a la variable dependiente del nivel explicativo.

La selección de la variable de supervisión depende del interés disciplinario. Por ejemplo, en un estudio sobre depresión y rendimiento académico, un psiquiatra situará la depresión como variable central, mientras que un educador priorizará el rendimiento académico. En ambos casos, lo esencial es reconocer que esta variable define la perspectiva del análisis y orienta la interpretación de las asociaciones halladas.

Las variables asociativas (auxiliares), denominadas también vinculantes, son aquellas que se ponen en relación con la variable de supervisión. Su función es contrastar, verificar y explorar si existe un vínculo estadísticamente significativo. Todas tienen igual jerarquía analítica, aunque su relevancia depende del grado de conexión que logren mostrar con la variable de supervisión.

Estas variables pueden abarcar dimensiones sociodemográficas, clínicas, conductuales o institucionales. Aunque no forman parte del objeto central de estudio, su valor radica en que permiten descubrir posibles factores asociados, generando así insumos empíricos para avanzar posteriormente hacia el nivel explicativo, donde se buscará determinar si tales asociaciones responden a relaciones de causalidad.

6. Variables analíticas en el nivel explicativo

A diferencia del nivel relacional, donde solo se constatan vínculos estadísticos, en este nivel se busca establecer qué factores producen un efecto determinado. Para ello, el análisis requiere un diseño riguroso, ya sea observacional o experimental, y se centra en dos categorías principales: la variable dependiente (variable de estudio) y las variables independientes (auxiliares).

La variable dependiente (variable de estudio) constituye el desenlace o fenómeno central que se intenta explicar. Puede tratarse de la incidencia de una enfermedad, la mortalidad hospitalaria, la deserción escolar o el rendimiento académico. Se requiere de una definición operacional, de modo que se garantice la validez de los resultados y la posibilidad de reproducir los hallazgos en distintos contextos y poblaciones.

La variable dependiente recibe distintas denominaciones según el campo del conocimiento: resultado, efecto, consecuencia o respuesta. En epidemiología se corresponde con el *outcome*; en psicología, con una conducta observable; y en educación, con un indicador de logro.

Las variables independientes (auxiliares) se dividen en observacionales y experimentales. En los estudios observacionales, se distinguen dos tipos: las probables causas, que son los factores de exposición propuestos como determinantes del desenlace, y las intervinientes, que abarcan variables de confusión, intermedias o de control, cuya función es ajustar y depurar la relación causal.

En los estudios experimentales, las independientes se dividen en manipuladas y no manipuladas. Las primeras son controladas deliberadamente por el investigador, como la dosis de un fármaco o la aplicación de un programa; las segundas son características como la edad o el sexo, que no pueden modificarse directamente. A efectos del análisis de datos, las probables causas en los observacionales equivalen a las manipuladas en los experimentales, mientras que las intervinientes corresponden a las no manipuladas.

7. Variables analíticas en el nivel predictivo

En el nivel predictivo, las variables analíticas se organizan en dos categorías: la variable endógena (variable de estudio) y las variables exógenas (auxiliares). A diferencia del nivel explicativo, cuyo propósito es demostrar causalidad, el nivel predictivo se orienta a anticipar resultados futuros, proyectando escenarios a partir de patrones previamente identificados.

La variable endógena (variable de estudio) llamada también variable a predecir, corresponde al desenlace que se desea anticipar o pronosticar. Es la continuidad de la variable dependiente, pero en este nivel su función ya no es explicar por qué ocurre un fenómeno, sino calcular la probabilidad de su ocurrencia o estimar su valor en el futuro.

Su definición operacional precisa es crítica, pues un error en su operacionalización compromete la validez del modelo predictivo. Además, debe medirse con instrumentos confiables y, de preferencia, estandarizados, ya que el desempeño del modelo dependerá en gran medida de la calidad de esta variable central.

Las variables exógenas (auxiliares) llamadas también predictoras, son aquellas que aportan información para estimar el comportamiento de la variable endógena. Representan los insumos del modelo. Su función es proveer las condiciones necesarias para anticipar con mayor precisión el desenlace.

Las exógenas no tienen el mismo valor jerárquico que la endógena, pero su selección es determinante para la solidez del modelo. Deben ser pertinentes y no redundantes ya que la validez del pronóstico depende tanto de la consistencia de estas variables auxiliares como del ajuste metodológico del modelo. De este modo, el nivel predictivo se constituye como un espacio donde las variables analíticas se convierten en herramientas de proyección y apoyo para la toma de decisiones basada en evidencia.

8. Variables analíticas en el nivel aplicativo

En el nivel aplicativo, las variables analíticas se dividen en dos tipos: la variable evaluativa (variable de estudio) y las variables de calibración o de control (auxiliares). A diferencia del nivel predictivo, que se orienta a anticipar escenarios, el nivel aplicativo centra su atención en la intervención directa sobre la realidad, midiendo los cambios alcanzados y asegurando que estos se consoliden en la práctica.

La variable evaluativa (variable de estudio) es el eje central de este nivel. Representa el resultado principal que se busca verificar tras la aplicación de una intervención, programa o estrategia. Su función es medir si los objetivos planteados se cumplieron y en qué magnitud. Puede adoptar naturaleza numérica, por ejemplo, niveles promedio de hemoglobina tras un programa de suplementación; o categórica, como la proporción de cumplimiento, éxito o fracaso de una acción concreta.

También recibe nombres como variable de impacto o de desempeño, porque refleja de manera directa los efectos de la intervención implementada. Su definición y medición rigurosa son indispensables, ya que de su precisión depende determinar si la acción aplicada produjo mejoras reales y sostenibles.

Las variables de calibración (auxiliares) acompañan a la variable evaluativa y permiten afinar la intervención. Su función es monitorear, ajustar y optimizar los procesos mediante los cuales se espera alcanzar el resultado principal. Se expresan como indicadores de proceso (cobertura, adherencia, tiempos de ejecución), de calidad (fidelidad en la implementación, exactitud en la medición) o de contexto (recursos disponibles, características poblacionales).

Estas variables, también llamadas variables de seguimiento, ajuste o monitoreo, garantizan la estabilidad y continuidad de la intervención. Al ofrecer información operativa y contextual, explican las condiciones bajo las cuales se obtuvieron los resultados evaluados.

Cuadro 18. Las variables analíticas según nivel de investigación

Nivel de investigación	Variable de estudio	Variables auxiliares
Exploratorio	Variable de estudio en construcción	Variables auxiliares cuando se requieran
Descriptivo	Variable de interés	Variables de caracterización
Relacional	Variable de supervisión	Variables asociativas
Explicativo	Variable dependiente	Variables independientes
Predictivo	Variable endógena (variable a predecir)	Variables exógenas (variables predictoras)
Aplicativo	Variable evaluativa	Variables de calibración

ACERCA DEL AUTOR

El Dr. José Supo es Médico Bioestadístico, Maestro en Ciencias Médicas y Doctor en Salud Pública. Director de www.bioestadistico.com, ha formado a miles de profesionales en metodología y análisis de datos. Autor de más de una docena de libros, entre ellos Metodología de la Investigación Científica, es referente en bioestadística aplicada e investigación científica en América Latina.

Programas de entrenamiento desarrollados por el autor:

1. Análisis de Datos Aplicado a la Investigación Científica
2. Seminarios de Metodología de la Investigación Científica
3. Validación de Instrumentos de Medición Documentales
4. Técnicas de Muestreo Estadístico en Investigación
5. Taller de tesis: Desarrollo del Proyecto e Informe Final
6. Análisis de Datos Categóricos y Variables Discretas
7. Análisis de la Causalidad con Diseños Experimentales
8. Soluciones de Análisis Predictivos Para la Investigación
9. Minería de Datos Para la Investigación Científica.
10. Control de la Calidad Para la Investigación Aplicada
11. Competencias Para Tutores, Jurados y Asesores de Tesis
12. Recursos Para la Redacción y Publicación Científica

MÁS SOBRE EL AUTOR

El Dr. José Supo es conferencista en métodos de investigación científica.

Libros y audiolibros publicados por el autor:

1. Cómo empezar una tesis
2. Cómo escribir una tesis
3. Cómo sustentar una tesis
4. Cómo ser un tutor de tesis
5. Cómo evaluar una tesis
6. Cómo asesorar una tesis
7. Taxonomía de la investigación
8. El propósito de la investigación
9. Las variables analíticas
10. Los objetivos del estudio
11. Cómo probar una hipótesis
12. Cómo elegir una muestra
13. Cómo validar un instrumento
14. Evaluación de pruebas diagnósticas
15. Técnicas de recolección de datos
16. Cómo se elige una prueba estadística
17. Cómo elegir una línea de investigación
18. El estudio de nivel exploratorio
19. El estudio de nivel descriptivo
20. El estudio de nivel relacional
21. El estudio de nivel explicativo
22. El estudio de nivel predictivo
23. El estudio de nivel aplicativo
24. Serie de mitos frecuentes

Sociedad Hispana de Investigadores Científicos

www.sincie.com

Descarga este y otros materiales en:
<https://bioestadistico.com/materiales>

BIBLIOGRAFÍA

1. Dahnke GL. La comunicación humana ciencia social. México: McGraw-Hill; 1986.
2. Tamayo y Tamayo M. El proceso de la investigación científica. México DF: Editorial Limusa; 2003.
3. Arias Odón FG. El proyecto de investigación: Introducción a la metodología científica. 6a ed. Caracas: Editorial Episteme; 2012.
4. Denzin NK, Lincoln YS. Handbook of Qualitative Research. 2a ed. Thousand Oaks, CA: Sage Publications; 2000.
5. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. 2a ed. Thousand Oaks, CA: Sage Publications; 2011.
6. Ander-Egg E. Aprender a investigar - Nociones básicas para la investigación social. 1a ed. Editorial Brujas, editor. Técnicas de investigación social. Córdoba: Lumen; 2011.
7. Carpio Valencia FE. Tipos y niveles de investigación. En: El proyecto de investigación para las ciencias empresariales. 1a ed. Arequipa: Universidad Nacional de San Agustín de Arequipa; 2018.
8. Alonso Trujillo J, Alonso Ricardez A. Aprendizaje estadístico basado en niveles de investigación. Revista Educación. 2022;46(1):1–16.
9. Supo Condori JA. Importancia del empleo de la bioestadística en las investigaciones biomédicas actuales. Revista Médico-Científica “Luz y Vida”. 2013;4:63–4.
10. Avellaneda Callirgos L, Morante Gamarra PC. Nivel investigativo. En: Investigación científica: Diseños y niveles de investigación. 1a ed. Bogotá: Instituto Latinoamericano de Altos Estudios; 2022.
11. Yunkor Romero YK, Ochoa Pachas JM. Validación interna de instrumentos de investigación en las ciencias sociales. Acta Jurídica Peruana. 2020;3(2):112–35.
12. Rodríguez Gómez R. Investigaciones cualitativas en salud pública publicadas en revistas biomédicas colombianas entre el 2011 y el 2021. Biomédica. 2023;43(1):69–82.

13. Tarrillo Saldaña O. Metodología de la investigación una mirada Global Ejemplos prácticos. Metodología de la investigación una mirada Global Ejemplos prácticos. CID-Centro de Investigación y Desarrollo; 2024.
14. Galeano ME. Diseño de Proyectos en la Investigación Cualitativa. Centro de Publicaciones Universidad EAFIT; 2020.
15. Ñaupas H. Metodología de la investigación Cuantitativa-Cualitativa y Redacción de la Tesis. Quinta. Ediciones de la U, editor. Bogotá; 2018.
16. Husserl E. Ideas relativas a una fenomenología pura. Halle: Niemeyer; 1913.
17. Heidegger M. Sein und Zeit [Ser y tiempo]. Tübingen: Niemeyer; 1927.
18. Gadamer HG. Wahrheit und Methode [Verdad y método]. Tübingen: Mohr Siebeck; 1960.
19. Sartre JP. L'Être et le Néant [El ser y la nada]. París: Gallimard; 1943.
20. Merleau-Ponty M. Phénoménologie de la perception [Fenomenología de la percepción]. París: Gallimard; 1945.
21. Schleiermacher F. Hermeneutics and Criticism and Other Writings. Cambridge: Cambridge University Press; 1998.
22. Dilthey W. The Formation of the Historical World in the Human Sciences. Princeton: Princeton University Press; 2002.
23. Habermas J. Knowledge and Human Interests. Boston: Beacon Press; 1971.
24. Ricoeur P. Freud and Philosophy: An Essay on Interpretation. New Haven: Yale University Press; 1970.
25. Nickles T. Heuristics and Discovery in Science. En: Gabbay DM, Thagard P, Woods J, editores. Handbook of the Philosophy of Science. Elsevier; 2007.
26. Lakatos I. Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge: Cambridge University Press; 1976.
27. Holyoak K, Morrison R. The Oxford Handbook of Thinking and Reasoning. Holyoak KJ, Morrison RG, editores. Oxford: Oxford University Press; 2012.
28. Pearl J. Heuristics: Intelligent Search Strategies for Computer Problem Solving. Reading, MA: Addison-Wesley; 1984.
29. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 3rd ed. Upper Saddle River: Prentice Hall; 2010.

30. Piaget J. *The Origins of Intelligence in Children*. New York: International Universities Press; 1952.
31. Vygotsky LS. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press; 1978.
32. Berger PL, Luckmann T. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books; 1966.
33. von Glasersfeld E. *Radical Constructivism: A Way of Knowing and Learning*. London: RoutledgeFalmer; 1995.
34. Delgado García G. Conceptos y metodología de la investigación histórica. *Rev Cub Salud Publica*. 2010;36(1):9–18.
35. Arias Odón FG. Investigación documental, investigación bibliométrica y revisiones. *Universidad Rafael Beloso Chacín*. 2023;31(22):9–28.
36. Vasconcelos SM. Rigor científico y ciencia abierta: desafíos éticos y metodológicos en la investigación cualitativa. *Scielo en perspectiva*. 2021;1(1):1.
37. Torres P YY. Análisis del discurso y análisis crítico en las ciencias de la salud 2024. *Innovación y gerencia*. 2024;(1):91–5.
38. Caballero-Rojas H. Perspectivas metodológicas del análisis del discurso aplicadas al estudio de los sistemas de salud. *Salud UIS*. 2023;55(1).
39. Urra-Medina E, Sandoval-Barrientos S. El análisis del discurso crítico en las investigaciones de salud. *Enfermería Universitaria*. 2018;15(2).
40. Cotán Fernández A. El método etnográfico como construcción de conocimiento: un análisis descriptivo sobre su uso y conceptualización en ciencias sociales. *Márgenes, Revista de Educación de la Universidad de Málaga*. 2020;1:83–103.
41. Remorini C. Etnografía y salud rural: trayectorias de investigación en los Valles Calchaquíes, Argentina. *Anthropologica*. 2020;38(44):267–96.
42. Suárez-Obando F, Gómez-Restrepo C. Aspectos éticos de la investigación etnográfica en salud. El papel del comité de ética de la investigación. *Pers Bioet*. 2017;21(2).
43. Ríos González CM. Investigación cualitativa en el contexto de la Salud Pública: actualización de conceptos. *Revista de Salud Pública del Paraguay*. 2024;14(1):51–8.

44. Guereca Torres R. Guía para la investigación cualitativa: etnografía, estudio de caso e historia de vida. Primera. Ciudad de México: Universidad Autónoma Metropolitana; 2016.
45. Glaser BG, Strauss AL. The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine Publishing Company; 1967.
46. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2a ed. Sage Publications, editor. Thousand Oaks, CA; 1998.
47. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. London: SAGE Publications; 2006.
48. Clarke AE. Situational Analysis: Grounded Theory after the Postmodern Turn. Thousand Oaks, CA: SAGE Publications; 2005.
49. Posada Hernández GJ. Elementos básicos de estadística descriptiva para el análisis de datos. Universidad Católica Luis Amigó; 2016.
50. Jordi Mas Elias. Análisis univariante. Barcelona; 2019.
51. Montanero Fernández J, Minuesa Abril C. Estadística básica para Ciencias de la Salud. Cáceres (España): Universidad de Extremadura; 2018.
52. Paliz Sánchez C del R. BIOESTADÍSTICA Introducción a la estadística en ciencias de la salud. Ecuador: Editorial BINARIO; 2024.
53. Espinoza Pajuelo LÁ. El nivel de investigación relacional en las ciencias sociales. Lima; 2021.
54. Aedo S, Pavlov D S, Ch FC. Riesgo relativo y Odds ratio ¿Qué son y cómo se interpretan? REV OBSTET GINECOL. 2010;5(1):51-4.
55. Ochoa Pachas JM, Romero Yunkor YK. El estudio descriptivo en la investigación científica. Acta Jurídica Peruana. 2019;2(2).
56. Flores Cebrián L. Análisis Estadístico Descriptivo. Lima; 2009.
57. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
58. Von Elm E, Altman DG, Egger M, Pocock SJ, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology

- (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4(10):e296.
59. López Himely AC. Ventajas del uso de estimaciones por intervalos para proporciones poblacionales. *Rev Méd Electrón.* 2018;40(3):876–81.
 60. González Landrián L. Intervalo de confianza para el porcentaje poblacional. *Revista Médica Electrónica.* 2020;42(3):1–8.
 61. Molina Arias M. Comparación de más de dos medias. *Análisis de la varianza. Evid Pediatr.* 2021;(17):11.
 62. Molina Arias M. Comparación de dos medias. Pruebas de la t de Student. *Evid Pediatr.* 2020;1(16):51.
 63. Ortega Páez E. Comparación de proporciones. Pruebas de χ^2 . *Evid Pediatr.* 2020;16:38.
 64. Ortega Páez E. Pruebas para muestras relacionadas. Variables cuantitativas. *Evid Pediatr.* 2022;18:41.
 65. Íñiguez Castro E. Estadística bivariada. *Estadística y tecnología de la información y comunicación.* Sevilla; 2020.
 66. Sagaró del Campo NM. Técnicas estadísticas para identificar posibles relaciones bivariadas. *Revista Cubana de Anestesiología y Reanimación.* 2020;19(2):1–3.
 67. Donis JH. Tipos de diseños de los estudios clínicos y epidemiológicos. *Avances en Biomedicina.* 2013;2(2):76–99.
 68. Moral Peláez I. Medidas de asociación. En: *Métodos Estadísticos Para Enfermería Nefrológica.* Madrid: Editorial Médica Panamericana; 2019.
 69. Osada J, Salvador-Carrillo J. Estudios “descriptivos correlacionales”: ¿término correcto? *Rev Med Chil.* 2021;149(9):1383–5.
 70. Bunge M. *La ciencia, su método y su filosofía.* 1a ed. Buenos Aires: Editorial Siglo XX; 1960.
 71. Kuhn TS. *La estructura de las revoluciones científicas.* México: Fondo de Cultura Económica; 1971.
 72. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med.* 1965;58(5):295–300.
 73. Gutiérrez Pulido H, De la Vara Salazar R. *Análisis y diseño de experimentos.*

- Segunda edición. McGraw-Hill Interamericana, editor. México, D.F.; 2008.
74. Mandeville P. Diseños experimentales. *Ciencia UANL*. 2012;15(57):151–5.
 75. Badii M, Castillo J, Rodríguez M, Wong A, Villalpando P. Diseños experimentales e investigación científica. *Innovaciones de Negocios*. 2007;4(2):283–330.
 76. Bacallao Gallestey J. Enfoques modernos del sesgo y la causalidad en la investigación epidemiológica. *Rev Cub Salud Publica*. 2012;38(5):686–701.
 77. Santos Buelga D, Dolores M. Aproximación al análisis de las relaciones causales en las ciencias de la salud. [Salamanca]: Universidad de Salamanca; 2018.
 78. Ochoa Sangrador C. Métodos de ajuste de sesgos. Análisis estratificado. *Evid Pediatr*. 2022;18:31–8.
 79. Quispe AM, Alvarez-Valdivia MG, Loli-Guevara S. Metodologías Cuantitativas 2: Sesgo de confusión y cómo controlar un confusor. *Rev cuerpo méd HNAAA*. 2020;13(2):205–12.
 80. Garcés D. Introducción al análisis multivariable. *Iatreia*. 2014;27(3):355–63.
 81. Espino Timón C. Análisis predictivo, técnicas y modelos utilizados y aplicaciones del mismo [Grado en Ingeniería Informática]. Universidad Oberta de Catalunya; 2017.
 82. Fernández Casal R. Métodos predictivos de aprendizaje estadístico. 2024.
 83. Carracedo P, Terrádez M. Minería de datos. Introducción y guía de estudio. Universidad Oberta de Catalunya; 2016.
 84. Beltrán Martínez B. Minería de datos. Benemérita Universidad Autónoma de Puebla, editor. 2021.
 85. Hernández V. Estudios epidemiológicos: tipos, diseño e interpretación. *Enfermedad Inflamatoria Intestinal al Día*. septiembre de 2017;16(3):98–105.
 86. Gómez Erik Cobo GY, Oliveres À. Análisis de supervivencia. *GH Continuada*. 2004;3(4):185–91.
 87. Martínez Pérez JA, Pérez Martínez PS. Análisis de supervivencia. *Medicina de Familia*. el 1 de julio de 2023;49(5):1–7.
 88. Rivera Rosales DD, Tejada DA. Fundamentos y aplicaciones del análisis de supervivencia para la investigación en salud. *Alerta, Revista científica del*

- Instituto Nacional de Salud. 2025;8(3):305–14.
89. Alonso Fernández AM. Introducción al Análisis de Series Temporales. 2007.
 90. Gallardo Pérez H de J, Gallardo J, Jhan P, Rojas P, Oscar S, Gallardo Pérez A. Modelación de series temporales en el sector productivo del Norte de Santander. 1a ed. Bogotá, Colombia: Universidad Francisco de Paula Santander; 2019.
 91. Manzano Patiño AP. Introducción a los modelos de ecuaciones estructurales. Investigación en Educación Médica. 2018;7(25):67–72.
 92. Russo Martínez N. Aplicaciones del análisis de supervivencia en la biotecnología [Grado en Biotecnología]. Universidad de Almería; 2021.
 93. Corres GA, Esteban A, Carlos García J, Zárata C. Análisis de series temporales. Revista Ingeniería Industrial. 2009;8(1):21–33.
 94. Bareño Silva J. Control estadístico de procesos en las organizaciones de salud. República de Colombia; 2023.
 95. Herrera RJ, Tomás A, Fontalvo Herrera J. Seis Sigma Métodos Estadísticos y Sus Aplicaciones. B-EUMED, editor. Buenos Aires, Argentina; 2000.
 96. Baudean M. Introducción a la investigación aplicada. Uruguay: Universidad ORT-FACS; 2015.
 97. Carrillo-Landazabal MS. Metodología DMAIC de Lean Seis Sigma: Una revisión en el contexto del ruido industrial. Ciencia Latina Revista Científica Multidisciplinar. 2022;6(2):3148–63.
 98. Gutiérrez Pulido H. Control estadístico de calidad y seis sigma. Tercera Edición. McGraw-Hill; 2013.
 99. Gironés Roig J. Metodologías y estándares CRISP-DM. Barcelona; 2013.
 100. Montalvo-García J. CRISP-DM/SMEs Metodología de Analítica de Datos para PYME. Medellín, Colombia: Universidad de Medellín; 2019.
 101. Carro Paz R. El Control Estadístico de Procesos. Universidad Nacional de Mar del Plata; 2015.

Reseña del libro

Niveles de investigación: Una clasificación emergente fundamentada en el análisis de datos para la investigación científica presenta un marco innovador que organiza la investigación en seis niveles: exploratorio, descriptivo, relacional, explicativo, predictivo y aplicativo.

El libro es resultado de una coautoría científica internacional, consolidada durante el Masterclass en Niveles de Investigación, realizado del 25 al 27 de julio en el *Hotel NH Collection Mexico City Reforma* (CDMX). En ese encuentro, investigadores y profesionales de diversos países aportaron sus experiencias, preguntas y reflexiones, integrando un conocimiento colectivo que dio origen a esta obra.

Más que un manual, es una guía práctica y motivadora que acompaña al investigador desde la primera idea hasta la aplicación de sus hallazgos para transformar la realidad.

Datos editoriales

Editorial: BIOESTADISTICO

Depósito Legal – Biblioteca Nacional del Perú: N° 2025-09491

ISBN: 978-612-99163-0-9

www.bioestadistico.com

Masterclass: Ciudad de México del 25 al 27 de julio del 2025

NIVELES DE INVESTIGACIÓN

Este libro surge del Masterclass, un evento donde investigadores hispanos convierten sus proyectos en publicaciones de alto impacto y reflejan la fuerza de una comunidad científica en acción.

Revive la experiencia del Masterclass y sé parte de la comunidad que la hizo posible.



Fotografías del Masterclass en CDMX 2025

José Supo

Es Médico de profesión, Profesor universitario e Investigador Científico. Presidente de SINCIE, impulsa la comunidad hispana de investigadores y dirige el Masterclass.



sincie

Sociedad Hispana de
Investigadores Científicos

Del Análisis de Datos a la Metodología de la Investigación Científica.

www.nivelesdeinvestigacion.com

ISBN: 978-612-99163-0-9



9 786129 916309