

LA METODOLOGÍA ESTADÍSTICA EN EL PROCESO DE INVESTIGACIÓN

COORDINADORES:

MARIO MIGUEL OJEDA RAMÍREZ

JUAN CARLOS ORTEGA GUERRERO

ARIEL ANTONIO LÓPEZ SALAS

La metodología estadística en el proceso de investigación

Coordinadores

Mario Miguel Ojeda Ramírez
Juan Carlos Ortega Guerrero
Ariel Antonio López Salas

Autoras y autores

Mario Miguel Ojeda Ramírez
Académico de la Facultad de Estadística e Informática
Miembro del Núcleo Académico Básico del Doctorado en
Innovación en Educación Superior

Juan Carlos Ortega Guerrero
Académico del Centro de Investigación en Innovación en
Educación Superior
Miembro del Núcleo Académico Básico del Doctorado en
Innovación en Educación Superior

Ariel Antonio López Salas
Egresado de la Facultad de Estadística e Informática
Auxiliar de Investigación de Mario Miguel Ojeda Ramírez

Rosbenraver López-Olivera López
Estudiante del Doctorado en Innovación
en Educación Superior

Joselin García Guzmán
Estudiante del Doctorado en Innovación
en Educación Superior

Leomar Mar Medina
Estudiante del Doctorado en Innovación
en Educación Superior

Estrella Thay-lli Armenta Courtois
Estudiante del Doctorado en Innovación
en Educación Superior

Juan Gerardo Domínguez Reyes
Estudiante del Doctorado en Innovación
en Educación Superior

Gerardo Yorhendi Ceballos Marín
Estudiante del Doctorado en Innovación
en Educación Superior

Víctor Hugo Ramírez Ramírez
Estudiante del Doctorado en Innovación
en Educación Superior

Carlos García Trujillo
Estudiante del Doctorado en Innovación
en Educación Superior

Exciani Aduy Alarcón Santamaría
Estudiante del Doctorado en Innovación
en Educación Superior

Jessica Badillo Guzmán
Estudiante del Doctorado en Innovación
en Educación Superior

Esta obra es una publicación de acceso abierto bajo una licencia de Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional. Para ver una copia de esta licencia, visita

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



© Mario Miguel Ojeda, Juan Carlos Ortega, Ariel Antonio López, 2024.

Sugerencia de referencia:

Ojeda, M. M., Ortega, J. C., y López, A. A. (Coords.). (2024).

La metodología estadística en el proceso de investigación.

Publicación de acceso abierto bajo licencia CC BY NC ND 4.0 Internacional.

La metodología estadística en el proceso de investigación

Este libro es una publicación de acceso abierto con los principios de *Creative Commons Attribution 4.0 International License*, que permite el uso, intercambio, adaptación, distribución y transmisión en cualquier medio o formato, siempre que dé el crédito apropiado al autor, origen y fuente del material gráfico. Si el uso del material excede el uso permitido por la normativa legal deberá tener permiso directamente del titular de los derechos de autor.

Contenido

Presentación.....	VIII
Introducción.....	XI
Investigación usando la metodología estadística.....	18
Métodos mixtos de investigación	36
El aprendizaje basado en proyectos en la educación superior .	53
La delimitación del problema, las preguntas y los objetivos de investigación.....	72
Diseño estadístico: tipo de estudios, unidades, variables y escalas de medición.....	88
Estructura y tratamiento preliminar de los datos.....	104
Análisis definitivos: análisis multivariantes y modelación estadística	133
Interpretación de resultados y elaboración de conclusiones..	172
Escritura del reporte y unificación editorial	184
Cultura estadística en la investigación	208

Presentación

La idea de este libro surgió a principios del 2022, en el marco del Seminario de Estudio IV del Doctorado en Innovación en Educación Superior, de la Universidad Veracruzana. Se pensó como un producto académico que fuera la base para las reflexiones y el trabajo de las futuras generaciones de este programa, pero, asimismo, que el proceso de diseño y elaboración permitiera que las y los estudiantes se entrenaran en las competencias de lectura y escritura académicas, realizando un trabajo colaborativo, cada quien a partir de su propio tema de investigación incursionando en la metodología estadística. Por supuesto, también se manejó la idea de que este libro llegara a tener la cobertura temática y la calidad suficientes que lo hicieran un material de consulta útil para quienes aborden la elaboración de proyectos de investigación factual, particularmente de proyectos que contemplen el uso de esta metodología, tanto en la fase de diseño como en el procesamiento de las evidencias empíricas, pero también en la elaboración de reportes de investigación y borradores de artículos con el acabado para enviarse a una revista arbitrada. Los coordinadores sabíamos que todo esto era un reto y que nos metería en una dinámica de trabajo muy intensa; y así fue.

Las primeras semanas las dedicamos a que las y los estudiantes leyeran a profundidad materiales relacionados con la metodología estadística en el proceso de investigación, tratando de que se comprometieran con el aprendizaje profundo; es decir, que revisaran los materiales que les proporcionamos, pero que también buscaran fuentes primarias, que hicieran revisiones críticas y que pudieran desarrollar

discursos suficientemente sólidos como para integrar un capítulo de un libro. Esta tarea realizada por el grupo mostró que contábamos con autoras y autores en potencia y que este método de estudio les motivaba de manera muy evidente. Por supuesto, las retroalimentaciones que los coordinadores dimos, así como las que entre ellas y ellos se hicieron, contribuyeron a esclarecer y delimitar el papel de la metodología estadística en el proceso de investigación, independientemente de si el énfasis del proyecto de investigación fuera cuantitativo o cualitativo. En este sentido, tratamos de precisar el rol del enfoque de Aprendizaje Basado en Proyectos (ABP) y la orientación que permiten los métodos mixtos para dar cabida a los principios y técnicas de la metodología estadística. Esta parte del Seminario concluyó con la presentación de la estructura tentativa para este libro y la asignación a cada quien de sus tareas de escritura y revisión.

Los capítulos fueron elaborados en borrador y retroalimentados por el primer coordinador de este libro; algunos fueron reescritos y por tal motivo en algunos casos hay coautores; esto se hizo así cuando las modificaciones, a juicio del primer coordinador, resultaron significativas. No obstante, todos los capítulos fueron leídos y aprobados por quienes los firman. Los coordinadores contamos con cartas de autorización de todos quienes aparecen en la lista de autoras y autores para que estos materiales pasasen al proceso editorial y que después se distribuyan libremente en la comunidad académica. Quienes aparecemos como coordinadores del libro queremos agradecer toda la disposición y el buen ánimo de las autoras y los autores de los capítulos. Asimismo, queremos agradecer a las doctoras Silvia Ivette Grappin Navarro y Sandra García Pérez, quienes

hicieron dictámenes del libro, lo que nos permitió que muchas partes y aspectos de la edición y la confección final mejoraran. No obstante, asumimos la responsabilidad que nos corresponde al haber coordinado este esfuerzo.

Xalapa de Enríquez, Ver., octubre del 2024.

Los coordinadores.

Introducción

La estadística es, primeramente, una disciplina científica. Como tal tiene una larga historia, que ya abarca varios siglos, aunque se dice que la estadística moderna nació en el siglo XX. Dado que usualmente se le pretende definir en frases cortas, se tiene de ella una visión limitada, cuando lo que se requiere es una comprensión amplia y justa. En este sentido, lo primero que hay que dejar claro es que la estadística es una ciencia, con fundamentos en la matemática (la llamada teoría estadística o estadística matemática), los cuales aquí no estudiaremos. ¿De qué se trata entonces este libro?

Lo primero que se quiere es llamar la atención sobre una incomprensión bastante generalizada: que se piense que la estadística son solamente los métodos que proporciona esta disciplina; es decir, se piensa que la estadística es lo que llamaremos aquí la metodología estadística. Una vez dicho esto, de lo que sí se trata este material es de dar una conceptualización e ilustrar los fundamentos sobre los cuáles se aplica la metodología estadística en el proceso de investigación. Así, en primera instancia, se quiere dejar establecido que cuando se dice que la estadística es la ciencia y el arte de diseñar y conducir procesos de inducción a partir de datos, los cuales deben ser pertinentes a un problema de déficit de conocimiento, se está en realidad hablando de los principios, técnicas y métodos que proporciona la estadística; es decir, se está hablando sólo de la

metodología estadística. En este sentido, la metodología estadística puede ser entendida como una tecnología para obtener conocimiento a partir de evidencias, a partir de datos. Entonces surge el proceso investigativo como el referente fundamental para entender la manera en que se aplica la metodología estadística.

El proceso de investigación toma como punto de partida la determinación de qué preguntas se quieren contestar. Esto implica un marco referencial: teorías y conocimiento previo sobre el fenómeno que se desea estudiar. Como se sabe, las preguntas se pueden contestar a través de los datos. Por ello, una vez que se definen las preguntas, se hacen precisiones sobre los datos, los cuales se conciben como mediciones de conceptos. Se precisa qué medir, cómo medir, sobre qué medir, cuántas unidades del colectivo medir, y qué unidades concretamente. Como se verá estos aspectos deben definirse con el mayor esscrúpulo si se está pensando en una investigación con un enfoque de mayor profundidad, como lo puede ser la el uso de los métodos de la estadística multivariante y la modelación estadística. Todo ello estrechamente formulado con base en lo que se quiere conocer, o sea, en el objetivo general de la investigación. En la misma medida, estas definiciones determinarán si es necesario obtener los datos (si es que ya están disponibles) o producirlos (si es que no existen). La estadística, a través de la metodología estadística, ofrece principios y herramientas para constatar que los datos sean pertinentes, significativos y confiables; porque los datos son la materia prima para obtener información —la cual tendrá que ser relativa a eso que queremos conocer; es decir, al déficit de conocimiento expresado en las preguntas de investigación—. Esta fase puede desarrollarse con

menores dificultades si se plantean preguntas concretas sobre el fenómeno o colectivo que se desea estudiar.

Para ilustrar esta parte a continuación se desarrollará un ejemplo. Un grupo de investigación sobre la lectura en una universidad —por ejemplo, la Universidad Veracruzana— desea saber qué tanto les gusta leer literatura a las y los estudiantes; así mismo, conocer qué tanto tiempo dedican a esta actividad; saber si prefieren leer literatura en libros impresos o en formato digital; conocer el concepto que tienen sobre la lectura de literatura; y saber su opinión sobre si la lectura en formato impreso les produce emociones diferentes que en el formato digital. Estas interrogantes son importantes en la era de la lectura digital; también interesa porque en muchas latitudes y ámbitos se están realizando este tipo de indagaciones. Hasta aquí, se sabe que si hacemos un cuestionario y lo aplicamos a las y los estudiantes se espera que sus opiniones sean diferentes dependiendo, por ejemplo, del sexo, de la carrera que estudian y del nivel de avance que tengan en su formación universitaria. Supongamos que la universidad en cuestión tiene una matrícula de alrededor de sesenta mil estudiantes distribuidos en carreras, a su vez agrupadas en áreas de conocimiento (exactas e ingenierías, humanidades y artes, biológicas y de la salud, económicas, sociales y administrativas). Es claro que aquí no se cuenta con los datos y se debe realizar el diseño de una encuesta por muestreo, la cual se puede aplicar a través de medios electrónicos. Hay que diseñar un cuestionario, probarlo y validarlo; así mismo hay que obtener un marco de muestreo y diseñar una muestra, que debe ser representativa por sexo, por área de conocimiento y nivel de avance en los estudios. El

marco referencial de la investigación contribuye a precisar estas definiciones.

Por otro lado, hoy en día, con las políticas de datos abiertos y acceso a la información pública, hay muchos estudios que sólo requieren saber dónde están los datos, cómo están dispuestos, cómo obtenerlos, etc. Aunque así sea, igualmente es indispensable contar con las preguntas concretas para saber cómo se van a organizar y a depurar estos datos. De esta manera, muchos estudios de economía y ciencias sociales hacen referencia a datos en fuentes secundarias; es decir, datos que son producidos y publicados por las oficinas o dependencias de estadística oficial. Esta es, sin duda, una gran área de oportunidad para aplicar la metodología estadística, sobre todo si se considera la amplia disponibilidad —también en forma abierta— de software estadístico. Ante este escenario la aplicación de la metodología estadística se está haciendo más frecuente, al grado de que conocer los principios, técnicas y métodos estadísticos, y sobre todo cómo trabajan en el proceso de investigación, se convierte en una imperiosa necesidad para las y los investigadores.

Una vez que los datos están disponibles y se ha garantizado su pertinencia y confiabilidad, como primer paso hay que realizar una labor de limpieza y comprensión de los datos. Muchas veces hay valores faltantes, errores de medición o de registro, etc., los cuales merecen un trabajo escrupuloso y la utilización de técnicas y procedimientos de cómputo, y métodos especiales que se han desarrollado para tal fin. No se debe subvalorar esta actividad puesto que es decisiva para contar con datos que puedan garantizar la producción del conocimiento que se requiere.

Usualmente la preparación de los datos implica mayor tiempo que los mismos análisis.

El análisis de datos se realiza a partir de 2 estrategias: primero una exploración inicial, que regularmente requiere técnicas gráficas y métodos numéricos, de resumen y extracción de estadísticas básicas, como frecuencias de categorías o valores, porcentajes, porcentajes acumulados, etc. Se deben identificar las variables clasificatorias y sus categorías, y a partir de éstas hacer segmentaciones y exploraciones comparativas en las variables, una por una, por pares, etc. Para ilustrar la idea se retoma el caso sobre la lectura de literatura en los estudiantes universitarios. Lo que interesa es estudiar el gusto, la dedicación, la preferencia y el concepto de lectura; las primeras variables se pueden definir de manera categórica, con categorías como: poco, regular y mucho; prefiero el impreso, el digital o cualquiera; sí, siento diferencias en las emociones, o no, etc. Aquí el concepto de lectura sería una variable de respuesta abierta, donde como dato para cada estudiante se tendrá una cadena de caracteres (se denomina un dato textual). Si ya se ha levantado la encuesta y limpiado el archivo de datos, se tendrá un registro para cada estudiante, donde se pueda identificar su sexo, su área de conocimiento, su nivel de avance en los estudios (que puede ser categorizado como inicial y avanzado), las categorías de respuesta de todas las variables categóricas y la respuesta abierta. Supongamos que se tiene una muestra de 1765 estudiantes (el tamaño y la distribución de la muestra se determina en el plan de muestreo); esto producirá una base de datos de 1765 renglones, con tantas columnas como variables, que en este caso serían ocho.

La segunda fase de la estrategia de análisis responde a preguntas específicas: por ejemplo, ¿hay diferencias en el gusto de la lectura de literatura por sexo?, ¿la hay por área de conocimiento?, ¿la hay por nivel de avance en los estudios?, ¿qué tan asociado está el gusto con la dedicación?, ¿qué tanto el gusto con la preferencia de formato?, ¿el concepto de lectura difiere por sexo?, ¿difiere por área de estudio?, ¿difiere por nivel de avance?, etc. Es claro que cada pregunta merece un conjunto de análisis que pasan por la elaboración de tablas de resumen de frecuencias y porcentajes (llamadas tablas de contingencia); se pueden elaborar gráficas comparativas de barras, y finalmente realizar pruebas de hipótesis usando métodos como el de la Ji-Cuadrada. El resultado de estos análisis daría elementos para contestar las preguntas y para discutir los resultados, comparándolos con los de otros estudios en otras latitudes, tomando en cuenta las teorías de sociología de la lectura, por ejemplo.

Hoy en día se dispone de un amplio arsenal de técnicas de minería de datos, incluyendo las de minería de textos, técnicas de clasificación, agrupamiento y modelación, que permitirían análisis más enriquecedores, pero incluso así se sigue la misma lógica: proveer respuestas a preguntas concretas. El proceso de aplicación de la metodología estadística no ha terminado con el análisis de los datos y la interpretación de los resultados; aún queda la fase final, que es la de la comunicación del conocimiento obtenido. Regularmente se debe hacer una presentación de lo relevante y significativo, muchas veces de manera ejecutiva, ante un grupo de tomadores de decisiones; otras veces en un informe que se publica o entrega a los interesados; algunas otras veces se debe redactar un

artículo científico. La buena presentación de los resultados requiere de conocimientos estadísticos, sobre todo del manejo de los principios de la disciplina y asimismo de una ética profesional, pero también de técnicas de representación gráfica y numérica, así como de competencias de diseño; por eso a veces se dice que la estadística es una ciencia, pero también un arte, lo que desde luego se refiere a la forma de llegar a construir juicios concluyentes y comunicar resultados. Actualmente los que se dedican al diseño y elaboración de infografías deben conocer estos aspectos; es decir, los equipos de trabajo que hacen este tipo de presentaciones para la comunicación y difusión de la información requieren también del conocimiento de la metodología estadística.

En este libro se encontrará un marco referencial amplio para entender la metodología estadística en el proceso de investigación. A su vez, consideramos que provee un esquema muy práctico para incursionar en la investigaciones que la requieran. Se espera que este libro sea un referente para investigadoras, investigadores y estudiantes, sobre todo de posgrado. Cualquier retroalimentación será bienvenida.

Los coordinadores.

Investigación usando la metodología estadística

MARIO MIGUEL OJEDA RAMÍREZ
ARIEL ANTONIO LÓPEZ SALAS
JOSELÍN GARCÍA GUZMÁN

Introducción

Al ser un tema extenso, en cuanto a sus aportaciones en diversas ciencias, es necesario en primer lugar plantear una definición de estadística; esto en razón de que esta disciplina también ha trascendido el contexto académico y está presente en prácticamente todos los ámbitos de la actividad humana. La estadística como una ciencia está compuesta por aspectos teóricos, metodológicos, así como de herramientas y métodos para su implementación en diversos ámbitos. La estadística, según Gil Flores (2003), es el conjunto de procedimientos sistemáticos implementados para llegar a datos cuantitativos y cualitativos recuperados sobre una situación real (a los cuales se les identifica como “resultados”); dichos procedimientos son utilizados para la recopilación, la organización, el análisis, la presentación y la interpretación de las evidencias obtenidas durante un proceso investigativo. Obviamente que este autor hace referencia sólo a una parte de la estadística, que es precisamente la llamada metodología estadística.

En este primer capítulo se ponen a relucir los beneficios y ventajas que la metodología estadística trae a la investigación si ésta se aplica correctamente. Enfatizando en que la investigación en cooperación con la estadística potencia la obtención de resultados enriquecedores, los cuáles incluso propician áreas de oportunidad para nuevas investigaciones.

Metodología estadística ¿Para qué?

La estadística puede ser descrita como una disciplina que brinda herramientas y métodos que permiten guiar apropiadamente, el diseño del estudio, la recolección de los datos y su exploración, para que posteriormente sean analizados con mayor profundidad mediante la variedad de técnicas que ofrece. Adicional a esto, la estadística también aporta elementos para la interpretación y presentación de los resultados obtenidos. Su objetivo es extraer la mayor cantidad de información a partir de un conjunto de datos. No obstante, se debe recalcar que todo esto es precisamente la metodología estadística, porque la estadística es mucho más que esto: cuenta con una teoría, una filosofía, y esos aspectos son los que determinan los “productos” que la estadística pone al servicio de los procesos de investigación, que son precisamente los métodos y técnicas estadísticas.

Así entonces, la metodología estadística es caracterizada esencialmente por la versatilidad que posee para involucrarse de forma útil en prácticamente cualquier ámbito de la vida humana. Su utilización se ha extendido hacia prácticamente todas las disciplinas: desde la biología, la ingeniería, las ciencias sociales y económicas, la agronomía, las ciencias administrativas,

la medicina, la psicología y la epidemiología, hasta áreas como la antropología, la pedagogía y en las ciencias políticas. Los recursos que ofrece la metodología estadística han contribuido de forma trascendental en la resolución de problemas, y cada vez se ha convertido en herramienta esencial para la investigación. En la actualidad, la necesidad de integrar la metodología estadística en el ámbito de la investigación se ha acentuado. Una necesidad que ha sido solventada por el creciente desarrollo de software especializado para aplicar herramientas estadísticas, haciendo con esto que los métodos de análisis sean más accesibles. Se ha logrado así que las aplicaciones de la metodología estadística tengan mayor alcance y sean más populares. Por eso, es posible concluir que la demanda de la metodología estadística se ha visto favorecida en las últimas décadas gracias a su creciente utilidad, pero también por el desarrollo de las TIC, y del software especializado, que ahora es también de uso libre.

Es preciso rescatar que los preceptos que se tienen acerca de la metodología estadística y su uso han ido evolucionando. A través de los años, se ha dejado de ver a estos métodos y técnicas como complicadas, que se limitan al uso de procedimientos matemáticos y de cálculo; por el contrario, hoy en día se le reconoce a esta metodología más por el valor que genera su aplicación en una variedad de contextos. Como lo menciona Hand (2008), el enfoque de la metodología estadística actualmente se dirige más hacia la investigación, hacia el proceso de obtención de conocimientos, que a los cálculos en sí, ya que los cálculos los dejamos para las computadoras. Sin embargo, esto no libra de la responsabilidad de comprender e interpretar los

resultados, ni de establecer la estrategia de análisis más apropiada para cada caso.

Actualmente, las aplicaciones que tiene la metodología estadística van más allá de la descripción y el resumen de datos; siendo la inferencia una de las principales áreas de interés, ya que ésta permite elaborar conclusiones acerca de una determinada población partiendo de un estudio realizado en un grupo menor conocido como muestra, cuyos elementos se suelen seleccionar aleatoriamente (Martínez Rizo, 2019). Concretamente, la metodología estadística se puede distinguir en dos niveles:

- a) **Descriptiva o deductiva**, cuando su finalidad es evidenciar aspectos característicos (promedios, patrones, variabilidad, tendencia, etc.), los cuales ayudan a efectuar comparaciones sin pretender sacar conclusiones generales. En ella se elaboran cuadros, gráficos, promedios, varianzas, proporciones y análisis de regresión, entre otras herramientas.
- b) **Analítica o inductiva**, la cual busca explicar el comportamiento de un conjunto de observaciones, tratando de probar la significación o validez de los resultados. A partir de ella se consiguen conclusiones que se extienden más allá de las muestras (Martínez Rizo, 2019).

Otra clasificación, de las muchas que existen, segmenta a los métodos de análisis estadístico en paramétricos y no paramétricos; la primera se encarga de hacer estimaciones y pruebas de hipótesis sobre variables con una distribución probabilística conocida (ej. normal, exponencial, binomial, etc.); la segunda no identifica *a priori* la distribución que siguen las variables

estudiadas, por lo que no se puede asumir que se ajustan a criterios o modelos preestablecidos; se dice que los métodos inferenciales derivados así son libres de distribución. De lo anterior se desprende que los mismos datos condicionan la distribución con la que se establecen las pruebas de hipótesis, ya que los procedimientos dependen de los supuestos distribucionales que se hagan (Gamboa Grauss, 2018).

A partir de la metodología estadística se puede obtener información útil sobre un proceso o fenómeno con un gran ahorro de tiempo y de recursos. Hoy en día, la metodología estadística es aceptada y solicitada con mayor confianza por la comunidad científica. Procesar, resumir, analizar, interpretar y comunicar correctamente la información que se obtiene a partir de los resultados de análisis estadísticos, se han vuelto actividades altamente valoradas para la investigación. Es por ello que la aplicación de los métodos estadísticos se ha vuelto más recurrente, no solo para la investigación, también para la gestión de protocolos, proyectos y programas.

Hablando de la investigación, Sampieri et al. (2014) la definen como un conjunto de procesos que se aplican al estudio de un fenómeno; estos procesos deben ser sistemáticos, críticos y empíricos. En tal sentido, la metodología estadística es un elemento integral para la formación de nuevo conocimiento y constituye un papel importante en la toma de decisiones informadas en los procesos de obtención y aplicación del conocimiento. Es decir, las herramientas que conforman el espectro de la metodología estadística (las cuáles se siguen expandiendo y varían entre áreas de conocimiento) parten de fundamentos matemáticos y de

la teoría estadística. Por consiguiente, dentro de la investigación, las decisiones basadas en la metodología estadística excluyen en gran medida los matices de subjetividad, los criterios arbitrarios y los juicios sesgados, los cuáles pueden comprometer la fiabilidad de la investigación. Por ello, se insiste en la necesidad por comprender cómo se aplica correctamente la metodología estadística. Este proceso no ha de ser desconocido para muchos. No obstante, como se verá en los demás capítulos, el trabajo en cada una de las fases de esta metodología requiere la adquisición de una serie de habilidades y conceptos para llevarlas a la práctica en la investigación.

Como se sabe, las técnicas para analizar datos continúan expandiéndose, otras ya existían sólo en ecuaciones e investigaciones que por el contexto en el que surgieron no era posible llevarlas a la práctica; pero todas, después, en complicidad con los avances del cómputo estadístico, se han hecho cotidianas. Esto refuerza la idea de que se requiere una constante actualización de aquellos usuarios que se empeñen en la investigación. Las llaves nuevas son las que abren nuevas puertas.

Ahora, en este contexto, si se aspira a ser pionero en el área de investigación, es de suma importancia que los descubrimientos sean organizados y sustentados desde un fundamento estadístico, haciendo uso de técnicas de novedosas y aprovechando al máximo las herramientas para diseñar, adquirir, manejar, analizar e inferir a partir de los datos. Trascender hacia el uso de la metodología estadística en el proceso de investigación amplifica las posibilidades de que las disciplinas contribuyan a la resolución de problemas complejos, se

...en el proceso de investigación

subsanen las lagunas en cuanto a los fenómenos de incertidumbre, se tomen decisiones basadas en evidencias, se rescate desde un nuevo enfoque la utilidad e importancia de la investigación en la construcción del conocimiento y nos encaminemos hacia una sociedad crítica, que utiliza los datos para describir el mundo que le rodea. Subamos a este tren.

La metodología estadística en el proceso de investigación

En la actualidad el proceso de investigación y la innovación científica se constituyen en elementos base para el avance de las sociedades. En este sentido, Bernal (2010) menciona que, de forma conjunta, esta nueva sociedad emergente, requiere entonces de individuos con una mayor capacitación para generar y apropiarse de conocimientos. Por lo tanto, los procesos y fases de las investigaciones deben ser cuidadosamente orquestados. La metodología estadística apoya en gran medida a dirigir apropiadamente el uso de las herramientas, métodos y técnicas.

Ahora bien, dentro del campo de la investigación educativa, la metodología estadística es considerada “un conjunto de métodos, técnicas y procedimientos para el manejo de datos, su ordenación, presentación, descripción, análisis e interpretación, la cual ha ayudado a resolver distintos problemas en el ámbito de la educación” (Gil Flores, 2003, p. 233), impulsando la adquisición de conocimientos sobre las realidades educativas, la toma de decisiones y la mejora de la práctica por parte de docentes y otros profesionales de la educación.

Asimismo, existe una tendencia de las organizaciones educativas públicas y privadas en cuanto a implicarse activamente en el desarrollo y difusión de recursos para sus fines. En este punto se resalta al Proyecto ALEA, el cual proporcionó instrumentos de apoyo para la enseñanza de la estadística para alumnos y profesores a niveles primaria y secundaria. Además, se han hecho mini censos escolares, para dar a conocer a los alumnos su utilidad, el tipo de información recogida y su procesamiento; sumado a ello, se busca aumentar el interés y colaboración de las familias y el público en general en la elaboración de dicho censo (Batanero y Godino, 2005).

A pesar de estos intentos, Gamboa Grauss (2018) ha puesto de relieve que muchos en la academia y el trabajo docente no cuentan con una formación adecuada en el uso de la metodología estadística, lo que ha provocado diversas dificultades en los estudios educativos, tales como:

- a) Poblaciones que no se corresponden con la naturaleza de sus unidades de análisis.
- b) Muestras que no son representativas de la población en función del contexto y la finalidad de la investigación.
- c) Muestreos intencionales por sobre los aleatorios, con posterior generalización acrítica de los resultados a la población.
- d) Escalas de medición que no se corresponden con los requisitos de sus indicadores.
- e) Conclusiones sobre la base de datos que no son de calidad.
- f) Generalizaciones sin determinar la incertidumbre.
- g) Toma de decisiones sobre la base de hechos aislados e insuficientes datos.

...en el proceso de investigación

- h) Selección arbitraria de pruebas estadísticas de validación de los resultados.
- i) Incoherencia entre objetivos de investigación, determinación de variables que proporcionan los datos y la fiabilidad de los instrumentos de medición (Gamboa Grauss, 2018, p.4).

Desde esta postura crítica, son diversos los errores que pueden cometerse en el proceso de indagación al momento de incorporar a la metodología estadística, como advierten Calderón García et al. (2019). Entre ellos destacan la falta de delimitación del tema de investigación; la lentitud en el avance por una excesiva rigurosidad cuantitativa; el error de no separar bien el conocimiento previo y ya consolidado; y la desorganización durante el proceso de diseño de instrumentos, recolección y codificación de datos, así como en el análisis y la interpretación de los mismos. Todas estas fallas acontecen sobre todo por la falta de preparación en metodología estadística; es decir, no se cuenta ni con los rudimentos del análisis multivariado, ni con directrices metodológicas elementales para la realización de sus estudios, como advierte duramente Martínez Rizo (2019).

Producto de estos errores, ha surgido la propuesta de que exista un programa de formación de profesionales de la educación para poder aplicar contenidos esenciales relativos a los diversos métodos estadísticos, a través de los cuales se les facilite la elaboración de análisis descriptivos e inferenciales sobre los datos conseguidos siguiendo protocolos correctos. Con ello se pretende potenciar las investigaciones educativas “desde la interpretación y la solución de problemas profesionales para arribar a conclusiones

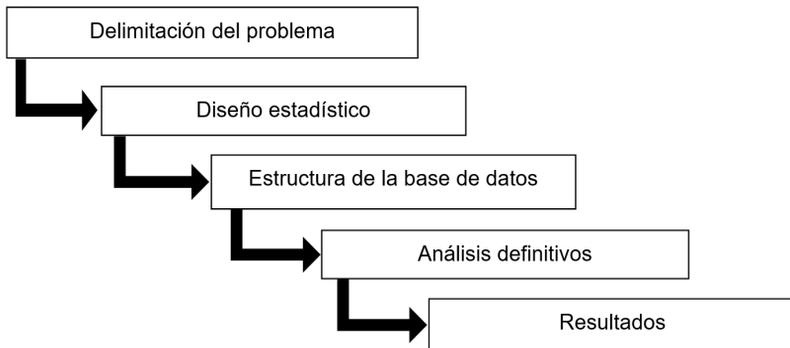
válidas y tomar decisiones razonables” (Gamboa Grauss, 2018, p. 4).

A pesar de todos los intentos por corregir posibles sesgos y valoraciones poco fiables, Ávila (2019) ha destacado que las herramientas de la estadística no deben ser empleadas “para enmascarar conscientemente la realidad, sino que sólo debe usarse para presentar determinadas perspectivas acerca del objeto estudiado” (p. 10). Por lo tanto, la idea de que la incorporación de sofisticados métodos y procedimientos estadísticos ofrece el rigor científico requerido en una investigación, no necesariamente es verdadera, pues la metodología estadística no puede suministrar autenticidad científica a un estudio que carece de los planteamientos, objetivos e hipótesis, que son los que determinan la calidad de la investigación. Vance y Smith (2019) destacan la importancia de delimitar y entender el problema del trabajo desde un enfoque orientado hacia la metodología estadística, indicando que la verdadera comprensión de éste permitirá tomar decisiones correctas en las siguientes fases de la investigación. Para esto proporcionan una serie de preguntas que dan información útil; entre ellas: ¿cuál es el problema principal?, ¿por qué es el problema importante o interesante?, ¿cómo será utilizada eventualmente la solución?, ¿cuáles son los datos que, potencialmente, permitirían resolver el problema?, ¿cómo fue o será la recolección de los datos? y ¿cuáles podrían ser las relaciones entre variables, para aquellas observadas y no observadas? Una guía más específica de cómo se lleva a cabo este proceso se presenta en el capítulo 4.

Cabe resaltar que ha sucedido que la selección de las metodologías estadísticas responde a ciertas modas,

por lo cual a veces se utilizan algunas técnicas de valoración que no corresponden al objetivo de estudio y pueden resultar inadecuadas para el procesamiento e interpretación de los datos. También ha ocurrido que el problema de investigación se somete al método estadístico que se desea probar para conseguir el reconocimiento de la comunidad científica, invirtiéndose así el orden epistemológico (López y Diez, 2017), pues las evidencias estadísticas sólo aportan correctamente cuando se combinan con el apropiado conocimiento teórico del tema abordado (Ávila, 2019). Por estas razones, es vital que se comprenda la relevancia de incorporar correctamente la metodología estadística, y más que sólo incorporarla, aplicarla apropiadamente dentro del proceso investigativo. Es necesario destacar que para implementarla es imprescindible un involucramiento profundo en el contexto del problema: que estudie fundamentos teóricos, que revise el conocimiento previo sobre el problema en cuestión y, sobre todo, que haga la delimitación del problema real y la traducción al problema estadístico, en el lenguaje de las dimensiones, variables, subvariables, etc. La compenetración con el tema, en sintonía con la metodología estadística, aporta una visión más clara acerca de qué camino tomar para guiar la investigación. La metodología estadística va estructurando e interviniendo con herramientas estadísticas en cada momento, logrando que los errores o sesgos sean minimizados. Dentro de cualquier investigación factual o fáctica; es decir, que considere datos para su análisis, el proceso estadístico puede ser fraccionado en 5 grandes momentos.

Figura 1. Esquema general



Nota: Elaboración propia.

Aunque se dará una descripción más específica de cada uno de estos momentos, aquí brevemente se indica a grandes rasgos de qué tratan:

1. **Delimitación del problema:** Se plantean las preguntas y objetivos que serán los ejes de la investigación.
2. **Diseño estadístico:** Se plantea el diseño de la investigación desde el enfoque estadístico, las variables, las escalas de medición, las unidades a analizar y cómo se realizará el proceso de muestreo o de obtención de los datos, de forma que se garantice la representatividad y la confiabilidad de la muestra.
3. **Estructura de la base de datos:** Se diagnostican los datos, evaluando su calidad, completitud y errores. En este mismo punto se utilizan también técnicas de análisis exploratorio sobre las variables de estudio.
4. **Análisis definitivos:** Se aplican las técnicas de análisis estadístico para dar respuesta a las preguntas que originaron. Es probable que durante este proceso,

...en el proceso de investigación

los productos obtenidos den origen a nuevas preguntas.

5. **Resultados:** Se describen e interpretan los resultados de los análisis realizados y se contrastan con los hallazgos de otras investigaciones. Se describen los elementos estadísticos y su significado, al mismo tiempo que se contextualiza y enriquecen estos resultados desde el marco del estudio.

Este esquema sintetiza cómo la metodología estadística contribuye al desarrollo de la investigación desde que se formulan las preguntas hasta que se elaboran las interpretaciones de los resultados. Evidentemente, dentro de cada momento, los conocimientos estadísticos deben ser aplicados correctamente, haciendo así que el desarrollo de la investigación fluya, al mismo tiempo que se construye un marco en el que cada aspecto descansa sobre el sustento de los principios y procedimientos de la metodología estadística. Al respecto, Ávila (2019) reconoce que el empleo de los métodos estadísticos ha trascendido los límites entre lo cuantitativo y cualitativo, estando presentes en estudios realizados desde perspectivas cualitativas, interpretativas o críticas.

A manera de conclusión

El papel que desempeña la metodología estadística en la sociedad va desde indagar información no obvia dentro de grandes conjuntos de datos a comprender el comportamiento de distintos fenómenos sociales y económicos, así como realizar descripciones o breves resúmenes numéricos, dándole un sentido y una interpretación a lo que observamos en la realidad.

La creciente especialización de las ciencias, las ingenierías y los procesos de las industrias han migrado a la utilización frecuente de métodos estadísticos (Martínez Rizo, 2019). Por otra parte, Vance y Smith (2019) al hablar del creciente trabajo entre estadísticos y científicos, señalan que el camino para maximizar el impacto y alcance de sus actividades está estrechamente relacionado con la colaboración entre estas partes, entendiendo que a veces se requiere de un profesional de la estadística en el equipo de investigación.

Cada vez existen más profesionales en el área de la investigación y especialistas que desean incorporar los métodos y técnicas estadísticas dentro de sus estudios e investigaciones. Como es conocido, hoy en día es indispensable que los resultados estén respaldados por evidencias y datos y que a su vez, hayan sido obtenidos siguiendo una metodología correcta. La proliferación de los datos y el acceso a las herramientas estadísticas, invitan a cualquiera que investigue a maximizar el impacto de sus resultados.

Lo que queda claro es que una vez que quien investiga se apropia del tema de estudio, será capaz de realizar diagnósticos de la situación, conociendo el marco que aquí se ha dado podrá aplicar correctamente la metodología estadística, por sí mismo o con el apoyo de un profesional de la estadística. Todo ello en armonía con el marco teórico al que está suscrito el problema. El estadístico, por su parte, puede tener múltiples ideas para analizar la información. No obstante, en ocasiones estas ideas pecan de la falta de sentido, o de ser poco relevantes en el contexto del área de investigación,

...en el proceso de investigación

misma que es conocida perfectamente por quien investiga.

Otro de los atractivos de la metodología estadística es su versatilidad de aplicación, lo cual ha permitido que los procesos de investigación en ciencias sociales se fortalezcan y consoliden a partir de los avances tecnológicos que han impulsado el tratamiento informático de los datos, tanto cualitativos como cuantitativos. La teoría de juegos, la ley de los grandes números, el análisis de series temporales, el muestreo, el cálculo de probabilidades, el control de calidad, las pruebas de hipótesis y los modelos estadísticos, son algunas de las herramientas que pueden aplicarse rutinariamente por profesionales en estudios sociales, económicos, lingüísticos y educativos para obtener información más precisa de los datos (Barreto Villanueva, 2012).

A lo largo de este capítulo se han abordado las cuantiosas e importantes aportaciones de la metodología estadística en la investigación (en especial en las áreas sociales y educativas), sobre todo al haberse abierto amplias perspectivas gracias a la posibilidad de acceder a bases bibliográficas en línea y de utilizar las plataformas digitales para obtener y trabajar cantidades enormes de datos, como bien refiere Martínez Rizo (2019).

Aunque la cercanía y la facilidad para el manejo de herramientas estadísticas es cada vez mayor, es aconsejable guiarse no sólo por la intuición y los objetivos, sino también por la asesoría especializada de un profesional de la estadística, como atinadamente sugieren Calderón García et al. (2019). De tal modo, si se trata de un tema teórico, dicho profesional puede

recomendar una técnica o modelo ya establecido para resolver el tipo de problema del que se trate. Por otra parte, si el tema elegido es aplicado, el especialista en métodos estadísticos puede ayudar a resolver un problema específico mediante la aplicación innovadora y eficiente de una o más técnicas o modelos apropiados. Bajo esta orientación, es mucho más factible que un científico social y educativo cuente con los elementos básicos para perfilar de mejor manera su investigación en el ámbito estadístico, evitar errores comunes (como la selección de la muestra o la incorporación de modelos incompatibles con el problema analizado), y fortalecer su interpretación y discusión de resultados.

Referencias

- Ayari Ávila, “La metodología estadística como herramienta básica para la investigación”, *Redieluz*, núm. 2, julio-diciembre de 2019, pp. 9-13.
<https://www.produccioncientificaluz.org/index.php/redieluz/article/view/32155/33360>
- Adán Barreto Villanueva, “El progreso de la estadística y su utilidad en la evaluación del desarrollo”, *Papeles de Población*, julio-septiembre de 2012, núm. 73, pp. 1-31.
- Carmen Batanero y Juan Godino, “Perspectivas de la educación estadística como área de investigación”, en *Líneas de investigación en didáctica de las matemáticas*, Ricardo Luengo (ed.), España, Badajoz: Universidad de Extremadura, 2005, pp. 203-226.

...en el proceso de investigación

César Bernal, *Metodología de la investigación*. Colombia, Prentice Hall, 2010.

Arturo Calderón García, Christian Bayes Rodríguez y Luis Valdivieso Serrano. *Guía de investigación en ciencias y en ingeniería*, Perú, Pontificia Universidad Católica de Perú (PUCP), 2019.

<https://cdn02.pucp.education/investigacion/2016/06/29210737/Guia-de-Investigacion-en-Estadistica.pdf>

Enrique Gamboa Grauss, “Estadística aplicada a la investigación educativa”, *Revista Dilemas Contemporáneos: Educación, Política y Valores*, núm. 2, octubre-enero 2018, pp. 1-32.

Javier Gil Flores, “La estadística en la investigación educativa”, *Revista de Investigación Educativa*, núm. 1, enero 2003, pp. 231-248.

<https://revistas.um.es/rie/article/view/99191/94791>

David Hand, *Statistics. A very short introduction*. Estados Unidos, Oxford University Press, 2008.

Abelardo López Domínguez y Fumero Díez, “Aproximación de la estadística a las ciencias sociales: una mirada crítica”, *Revista Cubana Educación Superior*, núm. 2, 2017, pp.148-156.

Felipe Martínez Rizo, *El nuevo oficio del investigador educativo. Una introducción metodológica*, México, Universidad Autónoma de Aguascalientes y Consejo Mexicano de Investigación Educativa (COMIE), 2019.

https://editorial.uaa.mx/docs/ex_nuevo_oficio_investigador_educativo.pdf

Roberto Sampieri., Carlos Collado y María del Pilar Baptista. *Metodología de la investigación*, México, Mc-GrawHill, 2014.

Eric Vance & Heather Smith, "The ASCCR frame for learning essential collaboration skills", *Journal of Statistics Education*, núm. 3, noviembre 2019, pp. 265-274.

<https://doi.org/10.1080/10691898.2019.1687370>

Métodos mixtos de investigación

ESTRELLA THAY-LLI ARMENTA COURTOIS
MARIO MIGUEL OJEDA RAMÍREZ

Introducción

Se conoce como métodos mixtos (MM) a aquellos que pueden combinar los métodos cuantitativos y cualitativos para obtener información dentro de un estudio. Hay una amplia literatura, de más de 50 años, como reconocen Forni y De Grande (2020), que muestra que la contraposición cualitativo-cuantitativo ha sido superada, por lo que estos paradigmas pueden articularse y complementarse.

De manera sostenida los esfuerzos de los especialistas en ciencias sociales se han centrado en la interconexión entre los juicios y narrativas cualitativas, por un lado, y los datos numéricos, por el otro. Esto bajo la noción del concepto de triangulación, surgido de la premisa de que una proposición comprobada a partir de dos o más procesos de medición independientes y complementarios reduce en gran medida la incertidumbre y el sesgo de su interpretación. De tal modo, la triangulación puede darse entre métodos de distintos enfoques, e incluso distintas corrientes teóricas (Forni y Grande, 2020). De esta concepción se desprendió la idea de crear metodologías mixtas. A ello sobrevino en los años noventa el planteamiento de dar primacía a las preguntas de investigación por sobre las preferencias metodológicas al momento de elegir las

estrategias de recolección de los datos y evidencias (Forni et al., 1993).

Sintetizando lo anterior, se puede mencionar que los MM permiten observar la realidad a partir de combinaciones de paradigmas metodológicos y de numerosas técnicas de indagación. Utilizan la concepción de un pragmatismo metodológico. En éste, el proceso de investigación —en sus principios de libertad para adicionar y mezclar nuevos elementos— encuentra semejanzas en cómo se cocina un platillo mediante una receta que puede variar y mejorarse en cada preparación. Es por ello que esta clase de métodos está al alcance de cualquier especialista con sólo incorporar determinadas habilidades técnicas (Cresswell, 2015).

El paradigma de que la metodología investigativa debe únicamente brotar de una sola raíz (cualitativa o cuantitativa) debe ser complementada por el reforzamiento de las nuevas alternativas, para llevar a cabo el proceso de investigación, en este caso, por los MM y así aprovechar sus bondades. El presente capítulo se dispone a enfatizar la idea de que, actualmente, cumplir satisfactoriamente con la complejidad de objetivos que implican una metodología mixta pueden lograrse haciendo uso de las herramientas que ofrece la estadística. Además, se construye un panorama general de los MM en la investigación educativa.

Los MM y la metodología estadística

Los MM implican la fusión de enfoques: cualitativo y cuantitativo. El uso de estos métodos permite beneficiarse de lo que ambos enfoques plantean. Esto suele ser sumamente útil para contestar o dar respuesta

a preguntas de investigación que son complejas. Como se sabe, los fenómenos de investigación pueden tener cierto nivel de profundidad. Sin embargo, si el estudio se limita a únicamente abordarlo desde una sola perspectiva (ya sea cualitativa o cuantitativa) es probable que el análisis no llegue a desbordar los límites que ir a profundidad implica.

Hamui-Sutton (2013) explica que los MM, al contener la perspectiva cuali y cuanti, permiten que el proceso de investigación pueda seguir su trayectoria a medida que ambas perspectivas van avanzando. Además, el panorama de la investigación logra expandirse considerablemente, permitiendo comprender el fenómeno con una mayor riqueza de elementos que terminan por otorgar mayor validez y novedad a los resultados.

Los retos que se plantea la investigación son trascendentales y, en esta medida, su desarrollo debe apoyarse en herramientas y recursos que permitan superar estos retos. En pleno siglo XXI debe reforzarse la cooperación entre los paradigmas de la investigación y los principios y métodos de la estadística. Una estrecha relación entre ambas partes da un paso más hacia plantearse propósitos de investigación más integrales, mismos que pueden cumplirse con ayuda de los MM, y apoyándose fuertemente en la metodología estadística, si consideramos que se cuenta con métodos de análisis de datos textuales, asociados a entrevistas, a discursos, etc.

Tipos de MM

Existen diversos modelos de los MM que pueden incorporarse al proceso investigativo. Estos modelos varían en función del rol que desempeñan los enfoques en la investigación (cualitativo y cuantitativo), ya que no siempre tendrán el mismo peso en el trabajo. En Hamui-Sutton (2013), se hace referencia a 4 modelos de MM. Aunque estos modelos surgen en el entorno de la educación médica, se adaptan bastante bien al contexto de investigación educativa:

1. **Modelo para desarrollar instrumentos:** Se basa en la generación de información cualitativa (a través del análisis de preguntas abiertas en los famosos grupos focales) para generar instrumentos cuantitativos (Cuestionarios o encuestas). Por lo tanto, el diseño del instrumento estará basado en la experiencia de los participantes y no solo en la visión de quién investiga. En este modelo, quién lleva a cabo la investigación llevará un análisis cualitativo de las respuestas de los participantes, para después construir un instrumento más preciso y balanceado. Los resultados del elemento cualitativo complementarán lo encontrado en el análisis de lo cuantitativo.
2. **Modelo explicativo:** Su fundamento es que, a través de un análisis cualitativo, se expliquen los resultados encontrados de la parte cuantitativa. En este caso, la indagación y exploración de los datos serán el punto de partida para llevar a cabo un estudio cualitativo a los participantes. Por ejemplo, se podría diseñar y aplicar un cuestionario en relación con los hábitos de lectura, en base al análisis de los resultados, se podría estructurar una

etapa cualitativa (mediante entrevistas, por ejemplo) para profundizar y elevar el nivel de comprensión de lo que se encontró.

3. Modelo de triangulación: En este tipo de modelos, se recolecta información cualitativa y cuantitativa al mismo tiempo. El cuestionario incorpora ambos elementos y se analizan en una sola etapa. Los resultados se presentan de forma conjunta. A diferencia de los casos anteriores, en donde, se muestra a uno de los enfoques como consecuencia del otro.
4. Modelo de transformación longitudinal: Estos MM, se caracteriza por recoger información de los participantes en diferentes momentos del estudio. Usualmente son utilizados para entrever la evolución de un grupo y evaluar un aspecto de interés (el rendimiento académico, el desempeño docente, por ejemplo). Los datos, en estos casos, se van construyendo sobre sí mismos. Puede ser vista como un ciclo, en el que un enfoque (cualitativo o cuantitativo) dará pautas para construir información del otro.

La distinción de estos 4 modelos es sumamente valiosa para la investigación educativa. Ya que la comprensión de los fenómenos dentro del terreno de las ciencias sociales exige de esta complementación entre lo cualitativo y lo cuantitativo. Estos modelos pueden servir de guía para establecer con claridad cómo se van a incorporar ambos enfoques en la investigación. Vale la pena destacar que la integración de los resultados requiere de una minuciosa reflexión. Es probable que, en ocasiones, los resultados de la parte cualitativa (por ejemplo) este en conflicto con los de la cuantitativa. Sin embargo, esto no debe ser visto como un punto en

contra; si no más bien como una oportunidad para abrir nuevas líneas de investigación al tratar de explicar las diferencias de los resultados, generando un nivel de comprensión más elevado.

Pereira (2011) rescata que los MM pueden organizarse en función del estatus de los enfoques que lo componen. Cuando el enfoque cualitativo y cuantitativo tienen el mismo grado de importancia; lo único que cambiará es el orden en el que se aplican. Pero, se debe tener en cuenta que siempre el diseño de los MM se ajustará a los objetivos que se planteen, de tal forma que se le terminará dando naturalmente mayor énfasis a alguno de los dos enfoques. El estatus implícitamente debe definirse, independientemente del modelo que adopte la investigación.

Ventajas de los MM

Como se sabe, las directrices que pueden guiar a una investigación son variadas. En este sentido, de acuerdo con las necesidades del proyecto, se escogerá un enfoque investigativo. Al momento de seleccionar el enfoque que mejor se acopla a la investigación, es necesaria hacer una valoración de las bondades o ventajas que dicho enfoque puede traer a la investigación. Por lo tanto, es fundamental comprender las circunstancias del problema y vincularlas con las cualidades del enfoque. Los MM cuentan con ventajas que los han posicionado como la elección idónea para proyectos e indagaciones alrededor de temas complejos.

1. Recogen y analizan rigurosamente tanto datos cualitativos como cuantitativos con base en las preguntas de investigación planteadas.

...en el proceso de investigación

2. Permiten integrar los dos tipos de datos recolectados mediante su combinación, su presentación secuenciada, o su inserción de unos en otros.
3. Sirven para dar prioridad a uno o bien, a ambos tipos de datos, todo esto depende del énfasis de cada investigación.
4. Se integran estos procedimientos en un solo estudio o en múltiples fases de un programa de estudio.
5. Los MM se enmarcan dentro de visiones filosóficas y teóricas del mundo.
6. Ayudan a combinar procedimientos en diseños de investigación específicos para dirigir el plan para la conducción de un estudio (Cresswell y Plano Clark, 2011).

De forma más específica, como lo mencionan Valdés y Almeida (2015), la narrativa y las opiniones se pueden utilizar para comprender lo que dicen los números; y viceversa, los números pueden ayudar a fundamentar lo que se encuentra en la narrativa. Además, los MM permiten responder a un mayor número de preguntas. Al fundamentarse en la combinación de enfoques, la respuesta a estas preguntas se presenta con un panorama más completo. En contraparte, es probable que, ante la inexperiencia, quién investiga encuentre dificultades para acoplar ambos enfoques a su problema, lo cual puede resultar en una limitante.

Derivándose de estas características, si los MM se conjuntan adecuadamente desde la reflexión profesional de alguien experimentado, pueden constituir una perspectiva metodológica única y coherente. Combinar las cualidades del enfoque cualitativo y del enfoque cuantitativo, logra alcanzar así un nivel elevado de comprensión del objeto investigativo, lo que incrementa

sus resultados y conduce a una notable aproximación a la complejidad de los hechos abordados (Moscoso, 2017). Por lo tanto, es vital que cada vez, con más frecuencia, la investigación educativa utilice como pilar a los MM para abordar fenómenos complejos y proponer soluciones a problemas en este ámbito. Pensando en las múltiples contribuciones que brindan estos métodos al trabajo de investigación y el desarrollo de la metodología estadística para analizar los datos y evidencias, no hay razones suficientes para no mirar hacia los MM.

Panorama de los MM

A pesar de todas las bondades y posibilidades que se derivan de los MM, su uso no se ha extendido. Contrario a lo que podría pensarse ante todo el panorama favorable ya descrito, la investigación de MM todavía ocupa una pequeña parte del trabajo publicado en muchas naciones subdesarrolladas. Por lo anterior se reconoce como indispensable su difusión y aplicación para diferentes proyectos (Ramírez Montoya y Lugo-Ocando, 2020). Como lo mencionan Valdés y Almeida (2015), cada vez que se alude a los MM en la investigación en las ciencias de la educación, no se llega a explicar con claridad cómo es que este enfoque se desarrolla a lo largo de la investigación. Por esta razón, no es fácil entrever cuáles son las aportaciones que pueden hacer estos métodos en este campo.

Cabe resaltar que los MM han cobrado especial valor en la investigación educativa, aunque existe un escaso desarrollo de la reflexión metodológica en la literatura especializada en español. Además, como lo describe Pereira (2011), en el ámbito de la investigación

educativa, se presenta con mayores bemoles una necesidad e interés por ir más allá de lo evidente en el análisis de los fenómenos. Por ello, es crucial que adquieran una mayor coherencia epistemológica que permita su integración en las ciencias sociales (Moscoso, 2017). Justamente es en la búsqueda de la congruencia y adaptación de los MM que Harris (2021) resalta que la investigación realizada en contextos no europeos/norteamericanos obliga a considerar la importancia de las filosofías de investigación no occidentales.

Ahondando en esta discusión cabe resaltar que a través de los MM alguien que investiga puede determinar, a su criterio, hasta qué punto los sujetos de estudio pueden participar en las actividades de recopilación de datos. De igual forma a contribuir en la compensación de riesgos y la obtención de recompensas. Aquí se antepone el concepto de posicionalidad, que es descrito como “la postura o posicionamiento de quien investiga en relación con los contextos sociales y políticos del estudio: la comunidad, la organización o el grupo participante” (Coghlan y Brydon Miller, 2014). Sobre esto se ha puntualizado que “la cosmovisión y los antecedentes de quien investiga afectan la forma en que él o ella construye el mundo” (Berger, 2015, p. 220). Por lo tanto, sus ideas propias influyen en cómo se interpretan los resultados y deben tomarse en cuenta en la etapa de análisis. La combinación de enfoques se debe ver reflejada desde que se plantea el problema, así como en la forma en la que se analizan y se explican las evidencias recabadas; es decir, los análisis de los datos y las evidencias, así como la interpretación de los resultados.

En este sentido, la importancia del contexto cultural influye en la posición de quien investiga en los tipos de preguntas que se pueden hacer en una investigación mixta. También en la forma en que se recopilan los datos y la calidad de éstos. Bajo esta presunción, se debe reflexionar sobre cuestiones tales como (Coghlan y Brydon Miller, 2014):

- a) ¿Mi posicionalidad y la de mi personal afectan el acceso a los datos necesarios?
- b) ¿Mi posicionalidad y la de mi personal de investigación tienen un impacto en la calidad de los datos?
- c) ¿Cómo debe actualizarse el enfoque metodológico ante distintos factores culturales?

Como consecuencia de lo anterior, hay varios autores (Creswell, 2015; Moscoso, 2017; Forni y Grande, 2020) que afirman que los MM requieren de una “normalización”. Esta incluye establecer un protocolo mínimo para asegurar la coherencia de una investigación y evitar caer en las críticas planteadas por Howe (1988) en torno a las diferencias entre tipos de datos, concepción del problema, las diferencias de análisis, la interpretación de resultados y las bases epistemológicas. Por tanto, este protocolo debe contener dos niveles de teorización:

- a) **Epistemología del método.** Conlleva una reflexión profunda sobre los paradigmas cuantitativos y cualitativos, así como de la naturaleza de los datos resultantes, generando articulaciones entre ambos tipos de datos.
- b) **Operacionalización del método.** Se denota explícitamente el empleo de los dos métodos

...en el proceso de investigación

(simultaneidad, secuencialidad), la existencia de un método prioritario, la función de la integración y las fases de la investigación donde cada uno interviene (preguntas, recolección de información, análisis e interpretación) (Bryman, 2006).

Las circunstancias deben justificar el uso de los MM en la investigación. La decisión debe ser acompañada de una profunda reflexión de cómo la combinación de enfoques presupone una solución óptima, y de cómo se llevará a cabo el proyecto. Por otra parte, Mendizábal (2018) explica que el reto principal que afronta la aplicación de una metodología mixta es: integrar ambos tipos de datos en una estrategia de análisis integral; de tal forma que estos enfoques estén en sincronía.

La operacionalización de los MM

Los MM están constituidos por las etapas que contienen habitualmente los problemas de investigación, que van desde el planteamiento de un problema hasta la elaboración de conclusiones; con la diferencia, de que los enfoques cualitativo y cuantitativo se entrelazan para dar sustento en cada una de ellas. La incorporación de los MM debe impulsarse desde la comprensión de su naturaleza, utilidades y su alcance. Esto se desarrolla, en gran medida, al entender cómo operan los enfoques cualitativos y cuantitativos de forma aislada (Valdés y Almeida, 2015). A pesar de esta riqueza en el diseño metodológico los MM presentan serios retos para las investigadoras e investigadores, sobre todo en cuanto a la incompatibilidad de los enfoques cualitativos y cuantitativos. Razón por la cual la elaboración o elección de un MM depende de los objetivos y las condiciones

particulares de una investigación, así como de los recursos con los que se cuenta y los obstáculos que deben superarse (Ramírez Montoya y Lugo-Ocando, 2020). Mendizábal (2018) explica que esta metodología debe utilizarse para confrontar problemas densos, o cuando las preguntas de investigación están impregnadas de cierto nivel de complejidad. Por otra parte, este mismo autor, alude a la necesidad de la formación de equipos interdisciplinarios de investigación con diversas perspectivas.

Con respecto al muestreo se ha resaltado que los MM facilitan la clasificación en acceso, la atención a la seguridad y las complicaciones derivadas del idioma. No obstante, hay desafíos para la disponibilidad de datos, el acceso a personal de investigación capacitado y la ética de quienes investigan. Actualmente, existe una gran variedad de estudios de caso que han utilizado los métodos mixtos debido a su operacionalización práctica. Resaltan sus beneficios del enfoque epistemológico en la comprensión de contextos locales, al mejorar la calidad de los resultados (Martel et al., 2022).

Sumándose a estas utilidades, se ha reconocido que los factores culturales pueden ser considerados y valorados en mayor medida a partir de los MM. Asimismo, los MM son muy útiles para conseguir la apertura de los participantes a temas polémicos. Dependen en gran medida de que quien investiga esté físicamente presente o no, lo que a su vez tiene implicaciones logísticas (Cilliers et al., 2015).

En lo concerniente a la disponibilidad de datos, a través de los MM se pueden plantear diversas entrevistas con expertos en un tema particular (funcionarios gubernamentales, empresarios, activistas,

etc.). Lo anterior se puede complementar con encuestas específicas para una determinada población (ej. universitarios, mujeres embarazadas, migrantes, etc.). En cuanto al ámbito educativo, se ha referido que en la literatura reciente de los estudios sociales y educativos hay un crecimiento notable de las publicaciones que integran tecnologías digitales y MM, lo que está acrecentando las posibilidades para investigar innovaciones académicas y socioeconómicas. A este respecto, de 311 artículos publicados de enero 2010 a enero 2020 en Web of Science (WoS) y Scopus, demuestran que los MM están cambiando las formas de realizar investigaciones al incorporar nuevas estrategias y procesos para lograr una comprensión holística, interdisciplinar y crítica de los acontecimientos más actuales (Ramírez Montoya y Lugo-Ocando, 2020).

A manera de conclusión

Se concuerda con Moscoso (2017) en el sentido de que los MM “pueden convertirse en un enemigo de la coherencia científica o en una poderosa herramienta de descripción, comprensión y explicación de los fenómenos sociales y educativos” (p. 647). Para lograr que puedan utilizarse a futuro para la identificación y la posterior resolución de diversas problemáticas, es indispensable tomar en cuenta su pluralidad, contextualizarlos a las circunstancias del estudio, así como sistematizarlos para su uso óptimo. Deben sustentarse en sólidos fundamentos teóricos y procedimentales. Recordando que los principios y las herramientas de la estadística ayudarán en gran medida a quien investiga en cada una de las etapas. Es sabido que la forma de recopilar e

integrar datos y evidencias de distintas fuentes y formas, se ha vuelto una tarea más sistematizada, que requiere del uso de estrategias de análisis que se implementan con la ayuda de software especializado, por lo que el acceso a un mayor número de programas estadísticos ha abierto caminos para hacer análisis de cualquier tipo de datos, incluidos los que producen las entrevistas, los diarios de campo y bitácoras, etc.

Desprendiéndose de lo antes señalado, los MM pueden ayudar a comprender mejor elementos diversos y complejos, como la pobreza, el bienestar, la calidad de vida, la migración, el desempleo, la marginación, la violencia de género, la vulnerabilidad ante los desastres y los conflictos armados o políticos, como describe Harris (2021). Estos temas requieren ser abordados de forma multidisciplinaria e interdisciplinaria por especialistas en desarrollo, de quienes formulan las políticas públicas y profesionales en diversos campos de las ciencias sociales. Para los científicos sociales resulta una enorme tarea pendiente la difusión, la consolidación y la aplicación extendida de los métodos mixtos en los próximos años.

Referencias

- Roni Berger, "Now I see, now I don't: Researcher's position and reflexivity in qualitative research", *Qualitative Research*, núm 15, 2015, pp. 219-234.
- Alan Bryman, "Integrating quantitative and qualitative research: how is it done?", *Qualitative Research*, núm. 6, febrero 2006, pp. 97-113.
- Jacobus Cilliers, Oeindrila Dube & Bilal Siddiqi, "The white-man effect: How foreigner presence affects

...en el proceso de investigación

behavior in experiments”, *Journal of Economic Behavior & Organization*, núm. 118, octubre 2015, pp. 397-414.

David Coghlan y Marie Brydon Miller, *The Sage encyclopedia of action research*, Reino Unido, SAGE Publications, 2014.

<https://doi.org/10.4135/9781446294406.n277>

John Cresswell, *A concise introduction to mixed methods research*, Reino Unido, SAGE Publications, 2015.

John Cresswell & Vicki Plano Clar, *Designing and conducting mixed methods research*, Estados Unidos, SAGE Publications, 2011.

Floreal Forni, María Antonia Gallart y Irene Vasilachis, *Métodos cualitativos II. La práctica de la investigación*, Argentina, Centro Editor de América Latina, 1993.

Pablo Forni y Pablo De Grande, “Triangulación y métodos mixtos en las ciencias sociales contemporáneas”, *Revista mexicana de sociología*, núm. 1, junio 2020, pp. 159-189.

http://www.scielo.org.mx/scielo.php?pid=S0188-25032020000100159&script=sci_arttext

Alicia Hamui-Sutton, “Un acercamiento de los métodos mixtos de investigación en educación médica”, *Investigación en Educación Médica*, núm. 8, junio 2013, pp. 211-216.

<https://www.scielo.org.mx/pdf/iem/v2n8/v2n8a6.pdf>

Jamelia Harris, “Mixed methods research in developing country contexts: Lessons from field research in six countries across Africa and The Caribbean”, *Journal of Mixed Methods Research*, núm. 2, abril

2022, pp. 165-182.

<https://doi.org/10.1177/15586898211032825>

Kenneth Howe, “Against the quantitative-qualitative incompatibility thesis or dogmas die hard”, *Educational Researcher*, núm. 8, noviembre 1988, pp. 10-16.

Rhiannon Martel, Shepherd Mathew & Felicity Goodyear-Smith, “He awa whiria—A “braided river”: An Indigenous Maori approach to mixed methods research”, *Journal of Mixed Methods Research*, núm. 1, enero 2022, pp. 17-33.

<https://doi.org/10.1177/1558689820984028>

Nora Mendizábal, “La osadía en la investigación: el uso de los métodos mixtos en las ciencias sociales”, *Universidad de Zulia*, núm. 2, febrero 2018, pp. 5-20.

<https://www.redalyc.org/journal/122/12260698001/html/>

Javier Nuñez Moscoso, “Los métodos mixtos en la investigación en educación: hacia un uso reflexivo”, *Cadernos de pesquisa*, núm. 164, junio 2017, pp. 632-649.

<https://www.scielo.br/j/cp/a/CWZs4ZzGJj95D7fK6VCBFxy/?format=pdf&lang=es>

Zulay Pereira, “Los diseños de método mixto en la investigación en educación: Una experiencia concreta”, *Revista electrónica Educare*, núm. 1, enero-junio 2011, pp. 15-29.

<https://www.redalyc.org/pdf/1941/194118804003.pdf>

María Soledad Ramírez Montoya y Jairo Lugo-Ocando, “Revisión sistemática de métodos mixtos en el

...en el proceso de investigación

marco de la innovación educativa”, *Revista Comunicar*, núm. 65, octubre 2010, pp. 9-20.

Esperanza Lucía Valdés y Lázaro Emilio Almeida,
“Algunas reflexiones sobre el enfoque mixto de la investigación pedagógica en el contexto cubano”,
Revista Universidad y Sociedad, núm. 2, abril 2015, pp. 23-29.

<http://scielo.sld.cu/pdf/rus/v7n1/rus03115.pdf>

El aprendizaje basado en proyectos en la educación superior

ROSBENRAVER LÓPEZ-OLIVERA LÓPEZ

Introducción

Los procesos de enseñanza-aprendizaje no son, ni deben, ser estáticos, rígidos e inflexibles. Por el contrario, deberían fomentar dinámicas de operación que incluyan diversas formas de acercarse al conocimiento y alcanzar los objetivos que promueve la educación como es el caso del Aprendizaje Basado en Proyectos (ABP), *Project Based Learning* (PBL), en inglés. El ABP es un método de aprendizaje flexible que permite a los estudiantes desarrollar habilidades y competencias, así como favorecer entornos de aprendizaje que fomentan la participación activa de los sujetos que lo emplean. Para conocer más sobre el ABP, se presenta aquí una aproximación a las características esenciales del método iniciando con una breve semblanza histórica. Posteriormente, se discuten los componentes clave o pilares del ABP, las etapas para implementar un proyecto y se discute el rol de la labor docente, el soporte que debe proporcionar a las y los estudiantes, así como el rol de los participantes y su activo involucramiento en la construcción de conocimientos bajo un enfoque de aprendizaje autónomo.

Semblanza histórica del ABP

El ABP es un método de aprendizaje que se centra en las y los estudiante para trabajar de manera activa en la planeación, elaboración y evaluación de un proyecto que proporcionará el desarrollo de competencias encaminadas a alcanzar el aprendizaje. Las ideas que fundamentan esta metodología provienen del constructivismo a partir de las tendencias de varios filósofos, educadores y psicólogos como Piaget, Vygotsky, Ausubel, Bruner y Dewey. La metodología considera principalmente que la elaboración del proyecto sea establecida a partir de las consideraciones previas y análisis, que se hacen desde la perspectiva docente. El fin es proporcionar a las y los estudiantes los conocimientos e instrumentos adecuados que les permitirán desarrollar las competencias necesarias para alcanzar el conocimiento y lograr los objetivos de aprendizaje (Estalayo et al., 2021).

William Kilpatrick es considerado el fundador de la metodología ABP, aunque su propuesta original dista un poco de lo que hoy día se hace en las instituciones educativas, por lo que su aproximación inicial fue entendida como un método de proyectos. En su propuesta, Kilpatrick consideraba que el aprendizaje no debía ser obligatorio sino más bien se debía fomentar el interés quien estudia para adquirir conocimientos (motivación intrínseca) y fomentar así la participación activa en proyectos con diferentes tareas y actividades en las que se destacaba la colaboración y la interacción social como características esenciales (Estalayo et al., 2021).

Sin embargo, las aportaciones de Kilpatrick al método de proyectos no son consideradas las primeras para el desarrollo del ABP. Los proyectos en sí son mucho más antiguos, destacando aquellos que se empleaban en la agricultura y otras actividades similares. En el ámbito pedagógico se le atribuyen muchas ideas al filósofo, psicólogo y educador estadounidense John Dewey, quien aportó la formación de elementos epistemológicos al método que fueron difundidos, en gran parte, gracias a otros autores como Stevenson, Wells y Krakowinzer. Por su parte, a Kilpatrick quien también fuera discípulo de Dewey, se le atribuye el haber presentado originalmente el método como una propuesta pedagógica en 1918, ofreciendo a la vez una explicación más precisa de lo que era el método de proyectos (Campos, 2017).

De acuerdo con Moallem et al. (2019), John Dewey realizó enormes aportaciones al campo educativo desde una perspectiva psicológica ya que intentó entender el aprendizaje como un proceso que se conecta a través de las experiencias de los individuos. Dewey consideraba que el aprender haciendo era provocado por un problema entendido como un fenómeno. Este requería ser explicado y de ese modo el sujeto empleaba todos sus conocimientos y experiencia para encontrar una solución al problema a través de la realización de diferentes tareas y acciones, que lo encaminaban a obtener un producto final. De hecho, sus contribuciones filosóficas sobre el aprendizaje también influyeron en el método de aprendizaje basado en problemas, que, dicho sea de paso, suele confundirse en ocasiones con el aprendizaje basado en proyectos.

En aquel entonces, el método ABP tuvo gran popularidad en varias disciplinas académicas como

historia, matemáticas, ciencias y otras asignaturas de carácter profesional, laboral y manual. Posterior a la década de los años de 1950, el método quedó un poco restringido a la educación técnica, hasta el surgimiento del constructivismo y las teorías que se centraban más en el aprendizaje que en la enseñanza. Más adelante, el método es repensado para adaptarse a las nuevas realidades educativas. Para la década de los 1990, el Instituto Buck para la Educación (BIE, por sus siglas en inglés), la Fundación Edutopia y el *Autodesk Foundation*, entre otros, difundieron ampliamente las ideas y técnicas del método ABP como una opción didáctica nueva en el campo de la educación (Campos, 2017; Ortiz et al., 2016).

No obstante, algunos autores argumentan que el ABP tiene una trayectoria mucho más amplia como se puede ver en la Tabla 1. En ella se ubican cinco etapas o fases de desarrollo mencionadas por Campos (2017).

Tabla 1. *Fases de desarrollo del ABP.*

1590	1765: Se inicia el trabajo en proyectos en las instituciones europeas de arquitectura.
1765	1880: Se convierte en método regular de enseñanza en otras áreas y se traslada a Estados Unidos.
1880	1915: Uso en escuelas técnicas y escuelas de educación básica.
1915	1965: Redefinición del método de proyectos en su pase de Estados Unidos a Europa.
1916	Presente: Redescubrimiento de la idea del proyecto y la tercera ola de su difusión internacional.

Nota: Fases de desarrollo del ABP de Campos (2017, p. 61).

Finalmente, se puede mencionar que en la etapa más reciente el ABP se ha visto potenciado enormemente por la incorporación de las TIC. Lo anterior permite que su difusión llegue a más ámbitos educativos y adquiera mayor popularidad y nuevas oportunidades de interpretación. El uso de dispositivos electrónicos, el Internet, las nuevas realidades socioeconómicas y educativas, así como los nuevos escenarios de empleo y de vida deben ser elementos suficientes para permitir que las metodologías de aprendizaje puedan adaptarse a las circunstancias actuales para optimizar el trabajo en los centros educativos.

El ABP como método didáctico y pedagógico

El ABP, al igual que otras metodologías, está sujeto a diferentes dinámicas y enfoques pedagógicos que determinan no sólo su estructura y aplicabilidad, sino también sus componentes básicos y pilares por los que se rigen. No se trata de un método de aprendizaje memorístico que tradicionalmente han impulsado otros métodos de enseñanza-aprendizaje. Por el contrario, se trata de un método que fomenta que las y los estudiantes pueda discutir, ensayar, corregir, elaborar, mostrar y buscar información, revisarla críticamente, reflexionar, entre otras cosas, para alcanzar el aprendizaje. En ese sentido, pone a trabajar a las y los estudiantes, eventualmente de manera colaborativa, en un proyecto que les ayudará a adquirir ciertas habilidades y aprender a aplicarlas. Para ello los proyectos en el ABP deben cumplir con al menos dos características: (a) tener sentido para los participantes para que encuentren la

voluntad de hacerlo bien, y (b) tener un propósito educativo que vaya de acuerdo con el tema o materia donde se está aplicando (Aritio et al., 2021). De acuerdo con la revisión de planteamientos realizada por este autor, existen varios elementos esenciales que deben considerarse como pilares para el uso del enfoque ABP; ver la Tabla 2, donde además se discuten brevemente las consideraciones de autores como Vergara (2016), Trujillo (2015), y Larmer y Mergendoller (2012). Una representación gráfica de los pilares del ABP se muestran en la Figura 1.

Tabla 2. *Pilares para el ABP.*

Contenidos significativos	Establecer una metodología de trabajo colaborativo en un proyecto contribuye a que los estudiantes logren un aprendizaje significativo y profundo. Para ello, es necesario que los estudiantes visualicen el proyecto como algo cercano a sus intereses y real a su entorno para que así sea significativo y relevante para ellos.
Manifestación espontánea del interés	El docente debe ser capaz de encontrar el potencial educativo de los estudiantes a partir de la motivación resultante de las interacciones, así como el momento que están viviendo para relacionarlo con los contenidos curriculares y analizar los resultados de aprendizaje obtenidos.
Creación de un escenario	El docente debe crear el contexto o escenario para iniciar el proyecto en caso de que éste no exista. Esto se puede lograr haciendo una revisión al tema que

se abordará en el proyecto y revisar las posibles formas de enfocarlos desde diferentes perspectivas.

Necesidad de saber

Se considera oportuno que al usar esta metodología se pueda comenzar estimulando los sentidos de los estudiantes con información sobre el proyecto que realizarán. Esto se puede lograr proyectando un video, realizando una lectura de algún texto o artículo relacionado, revisando alguna noticia o recibiendo una simple charla sobre el tema. El objetivo de estas acciones es incentivar al estudiante para que abra su curiosidad y surjan dudas o preguntas que le hagan querer saber más sobre el tema.

Una pregunta que dirija la investigación

Se requiere de una pregunta que guíe el proyecto, que sea estimulante, de respuesta abierta y centrada con lo que los estudiantes deben aprender. Si no existe una pregunta central que guíe el proyecto, los estudiantes podrían perderse fácilmente entre la información e incluso malinterpretar el sentido y fin que se persigue al realizar el proyecto.

Voz y voto para el alumnado

Para establecer la dinámica de trabajo en el ABP también es importante considerar lo que los alumnos piensan, sobre todo en el sentido del producto entregable. Es decir, se les pueden preguntar sus propuestas sobre lo que quieren llevar a cabo como proyecto y así la responsabilidad cae en (a) el entregable,

...en el proceso de investigación

(b) la distribución de tareas y actividades, y (c) el ajuste a los tiempos para lograr alcanzar el fin del proyecto.

Competencias del siglo XXI

El trabajo colaborativo es esencial en el ABP y son los estudiantes quienes se encargan de hacer la repartición de las tareas y responsabilidades que conlleva lograr el proyecto. La organización previa del profesor es importante para formar equipos y establecer dinámicas de trabajo previo.

La investigación lleva a innovación

Para lograr llegar a la pregunta guía que ayudará a realizar el proyecto es importante considerar varias preguntas. Al ir definiendo o acotando estas preguntas se logrará llegar a una definitiva. Es necesario entonces realizar búsquedas en libros, artículos, sitios web y demás recursos para llegar a algunas definiciones y conclusiones a partir del debate que suja de toda la información revisada.

Evaluación, retroalimentación y revisión

La supervisión por parte del docente es esencial para revisar los borradores o planes que los alumnos van elaborando a lo largo del proyecto. También se revisan las fuentes que van utilizando, y se monitorea de manera general y constructiva todo lo elaborado a lo largo del proyecto.

Presentación del producto final ante

Es importante considerar que los resultados del proyecto sean presentarse

una audiencia ante una audiencia real. Esto puede ser otra clase, con un grupo de padres de los estudiantes, o con otros compañeros, docentes. Esta actividad permitirá crear un entorno de reflexión sobre lo que los estudiantes han realizado en el proyecto, lo que podrían hacer posteriormente o sobre las competencias y habilidades que han adquirido o conocimientos aprendidos.

Figura 1. *Pilares del ABP.*



Nota: Pilares del ABP en Aritio et al. (2021, p. 12).

Fases o etapas del ABP

Existe una serie de fases para realizar los proyectos empleando el método ABP, las cuales se pueden entender como procesos que ayudan a implementar esta metodología. Aritio et al (2021) identifican 4 etapas: La primera es decidir el tema que se abordará en el proyecto, tomando en cuenta el contexto en el que se encuentra. La organización del proyecto se identifica como segunda fase; la labor de investigar y obtener información del tema del proyecto como tercera fase y por último la etapa de presentación del proyecto realizado.

La primera etapa se entiende como la presentación general del proyecto y un análisis de la situación inicial para conocer la realidad en que este se desarrollará. Requiere describir el contexto, determinar de qué recursos se disponen, detectar qué necesidades existen y finalmente, qué dificultades podrían enfrentarse. Es importante que el que coordina el proceso docente reflexione sobre el tipo de acompañamiento que dará y las estrategias que empleará para motivar a las y los estudiantes y estimular su participación activa. Es aquí donde el rol docente se diversifica al rol de tutoría con un enfoque multidisciplinario. A su vez se dirige a la gestión de los proyectos facilitando y promoviendo las competencias que han de ser adquiridas por los participantes (Aritio et al., 2021).

En la segunda fase, investigación acción, considera la oportunidad de comenzar a indagar y obtener datos y conocimientos relacionados con la problemática del proyecto. Es deseable que en esta etapa se pongan en marcha diversas estrategias como realizar lecturas,

elaborar síntesis, resumir y comparar ideas e información, entre otras. El objetivo aquí es desarrollar las habilidades que se desean adquirir. Pueden ser habilidades para la comunicación, para desarrollar el pensamiento creativo o de cualquier otro tipo que se desee. El trabajo colaborativo es tan importante para el aprendizaje social como para el aprendizaje autónomo. En la etapa de evaluación básicamente se busca poner enfoque en los resultados de aprendizaje que han logrado las y los estudiantes, valorar el producto final y su repercusión en el entorno. También es necesario orientar el trabajo hacia la reflexión de lo realizado al final del proyecto. Esta reflexión o cuarta etapa ayudará a identificar nuevas cuestiones que surjan a partir del trabajo hecho y generar incluso nuevas líneas temáticas para nuevos proyectos (Aritio et al., 2021).

Aunque esta secuencia didáctica de los pasos no es única y exclusiva, bien puede servir como una guía para garantizar una práctica similar donde se implemente el ABP. Por otro lado, autores como Imbernón et al. (2020) indican que las etapas más características del ABP corresponden a la selección del proyecto, su planificación, distribución de tareas, la elección de los recursos y materiales, así como la organización de la información, hasta llegar a las etapas finales donde se concluye el proyecto y se presenta. Destaca también la importancia de analizar y evaluar los productos realizados y los resultados obtenidos. Esto permitirá desarrollar otras propuestas didácticas para implementar la metodología del ABP.

Por otro lado, García y Pérez (2018) hacen mención de diferentes secuencias didácticas que pueden emplearse en el ABP. Llama la atención la propuesta que

viene de la Fundación Edutopia centrada en el diseño de unidades, llamadas las siete fases de un ciclo de proyecto: (a) introducción a la pregunta motriz; (b) introducción del reto final; (c) desarrollo de la pericia de la materia; (d) realización del reto final; (e) presentación del reto final; (f) respuesta a la pregunta motriz; y (g) evaluación sumativa. De igual forma hacen mención a una segunda propuesta realizada por el BIE, organización en Estados Unidos que proporciona apoyo y recursos a docentes para la implementación del ABP. Muestra a modo de guía una secuencia didáctica para el diseño y planificación de un proyecto concebido en cinco fases o actividades: (a) comenzar con el final en mente; (b) modelar la pregunta motriz; (c) planificar la evaluación; (d) trazar el mapa del proyecto; y (e) gestionar el proyecto. En el sitio web del BIE (<https://www.pblworks.org>) se pueden ubicar otros materiales y recursos sobre cómo implementar la metodología del ABP.

Llama la atención que en ambos casos se considera como pregunta motriz la parte inicial del proyecto, que bien puede servir para encaminar el trabajo colaborativo de las y los estudiantes. García y Pérez (2018), también proponen una secuencia propia para utilizar el ABP. Dividen el proceso en tres fases generales: (a) definición del proyecto, que considera los elementos base, principios o pilares del ABP que incluya no sólo los objetivos, sino también otro tipo de información que ayude a estructurar el proyecto; (b) soporte, en la que se preparan diversas actividades y espacios de aprendizaje por un lado y materiales que facilitarán el éxito del proyecto; y (c) organización, que puede ayudar a la planificación de las actividades de enseñanza-

aprendizaje que se realizarán durante el periodo académico.

El soporte a las y los estudiantes

La función docente en el ABP cambia en el momento en que se modifica el paradigma educativo. Se considera que quien la realiza deja de ser sólo quien transmite y reproduce de los conocimientos, así como la cultura y los valores tradicionales, para convertirse en alguien práctico-reflexivo, capaz de trabajar colaborativamente, fungiendo como promoción y facilitación del aprendizaje independiente y colaborativo de las y los estudiantes. Es prácticamente un agente cuya misión es fomentar el desarrollo de las competencias que son necesarias para la formación. Esta perspectiva requiere una participación activa de quien ejerce la función docente para ofrecer una adecuada tutoría o asesoría en las actividades del proyecto (Rekalde y García, 2015).

García y Pérez (2018) indican que inicialmente se debe recopilar información sobre las fortalezas y debilidades de las y los estudiantes. Esto permitirá determinar el tipo de soporte que se les dará de acuerdo con el proyecto planteado. Así, se podrá ajustar el apoyo a las tareas o actividades en las que ellas y ellos encuentren más dificultades. Posterior a realizar un análisis sobre las fortalezas y debilidades, se debe considerar la forma de motivarles. Para ello, los autores proponen utilizar el modelo de atención, relevancia, confianza y satisfacción (ARCS). Se deben considerar por igual estrategias que atraigan la atención de los estudiantes y hacer que vean la relevancia de su proyecto. En los casos en que el apoyo no sea suficiente,

se podrían enfrentar severas dificultades y en los casos en que el apoyo sea demasiado se podría comprometer la dimensión de autoaprendizaje que fomenta el ABP.

También es importante considerar el diseño del soporte tomando en cuenta los puntos críticos del proyecto, las tareas o fases que realizan las y los estudiantes donde muestran tener mayores dificultades o aquellos puntos angulares que son importantes para el éxito del proyecto. García y Pérez (2018) proponen utilizar el modelo de Jonassen (2011) que identifica tres tipos de apoyo: *Scaffolding*, *modelling* y *coaching*. El primero se centra en el entorno y naturaleza de la tarea intentando hacer llegar al estudiante más allá de sus capacidades a través de marcos de apoyo al aprendizaje. El segundo, distingue dos tipos: (a) *behavioral modelling*, que muestra al estudiante ejemplos de las tareas que ha de realizar y cómo realizarlas, y (b) *cognitive modelling*, que se centra en la toma de decisiones, el razonamiento y la argumentación que deben hacer los alumnos al mismo tiempo que desarrollan las tareas o pasos a cumplir en sus actividades. El último tipo de soporte, *coaching*, intenta instruir o amaestrar al estudiante, enseñarle o guiarle y dar acompañamiento para que pueda lograr sus objetivos de manera autónoma. En cualquiera de los casos y sin importar el modelo que se utilice, es necesario destacar la importancia que tiene el rol docente para guiar el trabajo y las actividades que han de realizar las y los estudiantes con el fin de alcanzar el aprendizaje significativo.

La autonomía y el aprendizaje en el ABP

Una característica destacada del ABP es sin duda el rol que deben asumir las y los estudiantes para alcanzar los objetivos de aprendizaje. Se trata de un rol completamente activo y participativo; deben involucrarse en un proceso de aprendizaje que implica realizar una investigación sistemática; asumir un rol activo para tomar decisiones sobre el proyecto que están realizando, la forma de acercarse e indagar en el tema seleccionado y la construcción del resultado final que deberá presentar. Es decir, adquiere un rol sumamente participativo que requerirá a lo largo de la resolución del proyecto (Sánchez, 2013).

Estas formas de enseñar y aprender bajo nuevos enfoques constructivistas promueven que sea la o el estudiante quien, individual o socialmente, construya el conocimiento aplicando procesos cognitivos básicos como seleccionar la información, retenerla, comprenderla, organizarla e integrarla a modo de autocontrol del aprendizaje, lo que comúnmente se identifica como autonomía del aprendizaje (Rodríguez, 2006).

De acuerdo con Rodríguez (2006), cuando se aplica el aprendizaje activo, la función docente debe crear un entorno donde se promueva pensar como expertos. El objetivo es aprender a aplicar los conocimientos en situaciones prácticas, prepararlos para el trabajo colaborativo, desarrollar técnicas de trabajo que les permitan cooperar con otros, y enseñarles a tomar decisiones en grupo. Básicamente propone facilitar las condiciones para un entorno de aprendizaje centrado en

...en el proceso de investigación

el estudiante, que fomente un pensamiento crítico, analítico, que permita su participación, se construya el conocimiento en un contexto sociocultural, y que promueva relaciones positivas entre los pares y quien ejerce la función docente.

Si se toman en cuenta estas consideraciones y se crean los espacios apropiados, las y los estudiantes serán capaces de continuar con su deseo de aprender, se encontrarán más motivados y podrán orientar su aprendizaje hacia el trabajo participativo y colaborativo; realizarán actividades encaminadas a concluir su proyecto y, así, adquirirán competencias bajo un enfoque de aprendizaje autónomo.

Conclusiones

El ABP es un enfoque que permite construir el conocimiento a partir de la realización de una serie de tareas y actividades que ayudarán a desarrollar las competencias necesarias para completar el proyecto. Es importante destacar que los proyectos deben tener un propósito educativo bien definido y tener sentido para las personas que lo elaboran. También es importante considerar los pilares del ABP a modo de metodología para su implementación, así como las fases que pueden o deben desarrollarse para alcanzar los objetivos planteados. Se habló del rol docente, que es facilitadora y guía en el proceso de aprendizaje. También del rol de los participantes, como agentes socialmente activos para colaborar con sus pares y establecer mecanismos de trabajo en grupo que les permitan construir su conocimiento. Sin duda, el cumplimiento de estos aspectos permitirá una experiencia de aprendizaje

significativa tanto para los participantes como para quien realiza la función docente.

Referencias

- Rebeca Artirio Solana, Laura Bergues Piazuelo, Tomás Cámara Pastor y María Eugenia Cárcamo, “Cuestiones clave para el trabajo en ABP: pilares, fases, beneficios y dificultades”, en *Iniciación al aprendizaje basado en proyectos. Claves para su implementación*, Alicia Pérez De Albéniz Iturriaga, Eduardo Fonseca Pedrero y Beatriz Luca Molina (coords.), España, Universidad de la Rioja, 2021, pp. 9-19.
- Agustín Campos, *Enfoques de enseñanza basados en el aprendizaje*, Colombia, Ediciones de la U, 2017.
<https://elibro.net/en/ereader/bibliotecauv/70303?page=58>
- Ander Estalayo, Sergio Gordillo, Adrián Iglesias y Mario López, “La historia del aprendizaje basado en proyectos (ABP)”, en *Iniciación al aprendizaje basado en proyectos*. Alicia Pérez De Albéniz Iturriaga, Eduardo Fonseca Pedrero y Beatriz Luca Molina (coords.), España, Universidad de la Rioja, 2021, pp. 5-8.
- Javier García y Jorge Enrique Pérez, “Aprendizaje basado en proyectos: método para el diseño de actividades”, *Revista Tecnología, Ciencia y Educación*, núm. 10, mayo-agosto 2018, pp. 37-63.
- Mahnaz Moallem, Woei Hung & Nada Dabbagh, *The wiley handbook of problem-based learning*. Estados Unidos, John Wiley & Sons, Inc., 2019.

...en el proceso de investigación

<https://0-ebookcentral-proquest-com.biblioteca-ils.tec.mx/lib/biblitestm/detail.action?docID=5660370>.

- Francisco Imbernón, Joan Rué y Max Turull, “La metodología (II): técnicas y estrategias de enseñanza”, en *Manual de docencia universitaria. Educación Universitaria*, Max Turull (coord.), España, UB Octaedro, 2020, pp. 201-229.
- David Jonassen, “Supporting problem solving in PBL”, *Interdisciplinary Journal of Problem-Based Learning*, núm. 2, septiembre 2011, pp. 95-119.
- John Larmer & John Mergendoller, “8 essentials for project-based learning”, *Educational Leadership*, núm. 1, marzo 2012, pp. 1-4.
- Tania Ortiz Cardenas, Regla Margarita Calderón y Dayana Travieso Valdés, *La enseñanza por proyectos y el aprendizaje basado en problemas (ABP): dos enfoques para la formación universitaria desde una perspectiva innovadora*. Cuba, Editorial Universitaria, 2016.
- <https://elibro.net/en/lc/bibliotecauv/titulos/71628>
- Itziar Rekalde Rodríguez y Jon García Vílchez, “El aprendizaje basado en proyectos: un constante desafío”, *Innovación educativa*, núm. 25, diciembre 2015, pp. 219-234.
- <https://doi.org/10.15304/ie.25.2304>
- Raquel Rodríguez González, “Diseño de entornos para el desarrollo de la autonomía en el aprendizaje”, *Aula abierta*, núm. 87, 2006, pp. 89-103.
- <https://redined.educacion.gob.es/xmlui/handle/11162/4504>
- José Sánchez, “Qué dicen los estudios sobre el aprendizaje basado en proyectos”, *Actualidad*

pedagógica, 2013. Consultado en:

https://www.estuaria.es/wp-content/uploads/2016/04/estudios_aprendizaje_basado_en_proyectos1.pdf

Fernando Trujillo, *Aprendizaje basado en proyectos. Infantil, primaria y secundaria*, España, Ministerio de Educación Cultura y Deporte, 2015.

Juan José Vergara, *Aprendo porque quiero. El Aprendizaje basado en proyectos (ABP), paso a paso*. Ediciones SM, 2016.

La delimitación del problema, las preguntas y los objetivos de investigación

JESSICA BADILLO GUZMÁN

Introducción

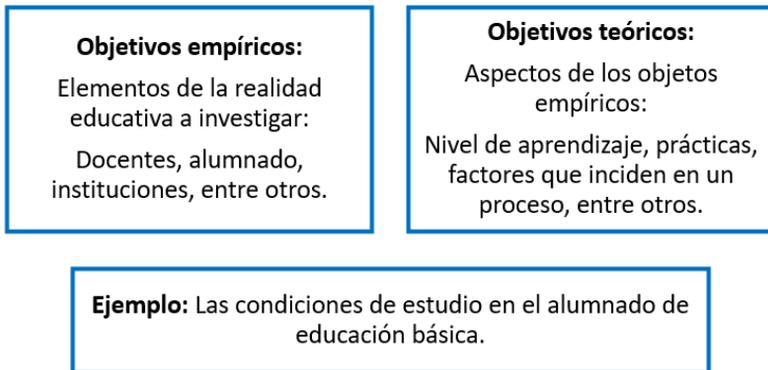
Definir el problema que abordará una investigación no es asunto sencillo. En algunos casos, quienes se aventuran a realizar un ejercicio de esta naturaleza tienen ya una idea del enfoque que quisieran seguir, pero no claridad en aquello que van a estudiar. Así, la construcción del objeto de estudio requiere delimitar el problema, las preguntas y objetivos a alcanzar, para posteriormente abundar sobre las motivaciones que dan lugar a la investigación, su pertinencia y relevancia, así como la revisión del estado del arte. Sólo entonces se estará en condiciones de avanzar en las cuestiones metodológicas y también teóricas. Partiendo de estas consideraciones, este capítulo aborda la delimitación del problema, preguntas y objetivos de investigación, constituyéndose como un referente para el proceso de construcción de tales apartados en un trabajo de investigación.

La delimitación del problema

¿Sobre qué investigar? Esta es la pregunta que apremia a las y los investigadores noveles con frecuencia, principalmente cuando no se tiene un campo de estudio

o una línea de investigación definida. Diferentes fuentes brindan elementos para responderla. Martínez Rizo (2019) explica que en primer lugar es necesario identificar un tema de interés, el cual se formula a partir de objetos empíricos u objetos teóricos o construidos teóricamente, tradicionalmente conocidos como objetos materiales y formales. La Figura 2 muestra un ejemplo para la delimitación del problema en el campo educativo.

Figura 2. *Objetos empíricos y teóricos en la delimitación del problema.*



Nota: Construcción propia con base en Martínez Rizo (2019, p. 39).

Arribar al tema de interés puede ser mucho más complejo de lo que hasta aquí se ha descrito. Por ello, es recomendable que, quienes estén iniciando el planteamiento de una investigación, puedan:

- a) Revisar publicaciones académicas sobre el tema de interés. Leer artículos de investigación, por ejemplo, puede ayudar a generar ideas y a precisarlas, a partir de los trabajos y hallazgos que otros han realizado. Esto también puede ayudar a identificar posibilidades

metodológicas que será necesario resolver en un momento posterior.

- b) Consultar los estados del conocimiento del campo disciplinar en el que se ubica la investigación. Por ejemplo, en el caso de la investigación educativa en México, el Consejo Mexicano de Investigación Educativa edita cada 10 años los estados del conocimiento relativos a este campo; quienes estudian cuestiones de educación, pueden encontrar allí referentes importantes para definir su problema de investigación, así como informes de metodologías y enfoques teóricos que fundamentan los trabajos.
- c) Cuando se trata de investigaciones en el marco de estudios de posgrado, se recomienda revisar las líneas de generación y aplicación del conocimiento del programa educativo, de modo que la investigación propuesta se ubique en una de ellas y abone a su construcción o consolidación.

Una vez que se ha definido un tema de interés, es necesario precisar el problema. En esta etapa, quién investiga puede apoyarse en la familiaridad que tiene con el tema, la complejidad del mismo, así como la revisión de investigaciones previas (Ojeda Ramírez, et al., 2011). Espinoza Freire (2018) apunta otras fuentes para identificar problemas de investigación. Entre ellas figuran el diálogo con otras investigadoras o investigadores, la experiencia y los registros de investigaciones al respecto. A decir de Martínez Rizo (2019) este proceso requiere de dos tipos de delimitación: empírica y teórica, esto es, de menor y mayor complejidad, respectivamente. La primera tiene que ver con los objetos o sujetos de estudio y sus características. La segunda requiere de la determinación de aquello que se estudiará de los objetos o sujetos.

Espinoza Freire (2018) señala que en el inicio del proceso investigativo se debe identificar una situación que todavía no tiene solución, concretarla y someterla a una valoración crítica, para estudiarla sistemáticamente. Así, además de establecer qué se va a investigar, debe esclarecerse con quién o quiénes, en dónde y, en algunos casos, en qué periodo de tiempo.

Arias Castrillón (2020) apunta que un problema de investigación debe contener: una descripción del fenómeno (¿qué sucede en el fragmento de realidad que se va a investigar?), una descripción del desequilibrio (¿por qué es un problema?), una descripción espacio temporal (justificar el contexto y de la duración de la investigación; situar históricamente el problema) y una descripción de los sujetos de investigación (características de la población de estudio y justificación de esta). Este autor reconoce que a partir de estas descripciones se podrá delimitar el objeto de la investigación y construir la pregunta que guiará el proceso investigativo. Lo anterior implica ubicar el problema en una realidad concreta. En este orden de ideas, el ejemplo descrito en la Figura 2, podría completarse como se muestra en la Figura 3.

...en el proceso de investigación

Figura 3. Ejemplo de delimitación del problema de investigación.

<p>¿Qué se va a investigar? Las condiciones de estudio.</p> <p>¿Con quién o quiénes? Alumnado de primer grado de telesecundaria.</p> <p>¿Dónde? Escuela telesecundaria Justo Sierra Méndez.</p> <p>¿En qué período de tiempo? Durante el primer año de la pandemia por COVID-19</p> <p>Problema de investigación:</p> <p>Las condiciones de estudio del alumnado de primer grado de la Telesecundaria Justo Sierra Méndez durante el primer año de la pandemia por COVID-19.</p>

Nota: Elaboración propia.

Además de estar delimitado, el problema debe cubrir otras características, tales como: estar formulado de manera precisa y clara; no dar cabida a ambigüedades; ser factible de investigarse; responder a una necesidad de conocimiento; poseer contrastabilidad empírica; ser novedoso; ser pertinente; tener relevancia teórica y metodológica; y poner en relación dos o más variables.

Una vez que se ha delimitado el problema se está en condiciones de elaborar las preguntas y los objetivos de investigación. Entre éstos deberá cumplirse el criterio de congruencia interna, que refiere a la relación directa entre dichos elementos.

Las preguntas de investigación

Según Garrocho y Amaury (2012) (como se citó en Ramos Galarza, 2016) una vez que quién investiga tiene claro el problema que quiere resolver, puede plantear interrogantes de investigación, lo que será la base para la elaboración de los objetivos y la elección de una

estrategia —teórico metodológica— para darle respuesta. Patiño y Carvalho Ferreira (2016) señalan que la definición de las preguntas de investigación impacta además en cuestiones como las hipótesis, el diseño del estudio, las variables y el diseño de la investigación, expresado en un protocolo. Por lo tanto, las preguntas de investigación constituyen los ejes sobre los que se orientará el estudio. Abarcan desde la formulación de objetivos hasta la construcción del diseño metodológico, el análisis de los resultados y las conclusiones a las que se llegue una vez realizada la investigación. Ramos Galarza (2016) refiere un conjunto de características que las preguntas de investigación deben reunir: actualidad, aporte al conocimiento, viabilidad, factibilidad, pertinencia, precisión, interés y ética. La Tabla 3 muestra una descripción de estas características.

Tabla 3. *Características de una pregunta de investigación.*

Elemento	Descripción
Actualidad	Original, novedosa y actual, evitando investigar cuestiones ya resueltas sobre un problema de investigación.
Aporte al conocimiento	Considerar el nivel en que se realiza la investigación. Licenciatura: aplica un conocimiento adquirido en la carrera. Doctorado: genera conocimiento nuevo en una línea de investigación.
Viabilidad	Se cuenta con los recursos para realizar la investigación y responder a las preguntas.
Factible	Reconocer si se cuenta con las condiciones, recursos e incluso participantes para llevar a cabo la investigación.
Pertinencia	Plantea la resolución de situaciones o problemas reales, correspondientes al campo

Elemento	Descripción
Precisión	disciplinar. Ubicación en un espacio de tiempo y contexto determinados, así como la definición de una población de estudio.
Interés	Ser del interés de quien efectúa la investigación.
Ética	Tener presentes parámetros éticos en cuanto al abordaje, salvaguardando la integridad de los participantes.

Nota: Elaboración propia con base en Ramos Galarza (2016, p. 26).

En cuanto a la cantidad de preguntas a elaborar, se recomienda contar con una pregunta general de investigación. De ella se desprenderán preguntas particulares. La pregunta general tendrá relación con la o las variables de estudio, mientras que las preguntas particulares harán referencia a los indicadores (cuando se trata de una investigación cuantitativa), o bien, a las categorías y subcategorías de análisis, respectivamente (cuando se trata de una investigación de corte cualitativo). Así, no hay una cantidad única o estandarizada de preguntas a formular. Esto depende del objeto de estudio, su alcance y nivel de profundidad. Iglesias y Cortez (2016, como se citaron en Espinoza Freire, 2018), subrayan que las preguntas pueden ser modificadas a lo largo del estudio, pudiendo reelaborarse e incluso agregar nuevas a partir de los hallazgos.

Otros elementos que pueden dar lugar a cambios en las interrogantes son el marco teórico y el contexto en donde se investiga. Mientras se construye el marco teórico pueden encontrarse nuevos aspectos del objeto

de estudio que se pueden considerar relevantes y, por tanto, incluirse en las preguntas. En el mismo sentido, cuando se indaga sobre el contexto es posible que se reconozcan en éste cuestiones que, por ser específicas, no se habían tomado en cuenta al elaborar las preguntas de la investigación, pero que pueden integrarse si se valora como pertinente hacerlo.

Ramos Galarza (2016) distingue entre las preguntas de investigación de un estudio cualitativo y uno cuantitativo a partir de los paradigmas que sustentan a cada uno. Esto es, el constructivismo y la teoría crítica en el cualitativo, y el positivista y postpositivista en el cuantitativo. El primero dará lugar a procesos de interpretación en donde las características particulares de los sujetos u objetos son importantes; el segundo involucrará procesos de medición estadística y generalizaciones. Para el enfoque mixto, más pertinente en la actualidad, se recomienda revisar, en este mismo volumen, el capítulo correspondiente.

De este modo, se pone en relación la formulación de las preguntas con el enfoque de la investigación, distinguiendo entre su carácter medible o interpretativo, según sea el énfasis cuantitativo o cualitativo. Sea uno u otro, hay algunos aspectos que deben considerarse al construir las preguntas de investigación:

1. Evitar preguntas con respuesta dicotómica. Si se plantea una pregunta que se responde con sí o no, no se abre el camino de la investigación. Ejemplo: ¿Es posible medir el nivel de aprendizaje del alumnado? Esta pregunta se responde con sí o no, sin dar lugar a un desarrollo teórico metodológico. Preguntas de investigación que inician con cómo, por qué, qué, cuál,

...en el proceso de investigación

darán lugar a un proceso de investigación (Ramos Galarza, 2016).

2. La pregunta debe referir a una población de estudio y estar contextualizada. Ejemplo: ¿Cuál es el nivel de apropiación tecnológica de las y los estudiantes del Telebachillerato Las Trancas? En esta interrogante se define una población y un contexto institucional específico, lo que delimita los sujetos y el objeto de estudio, además de que reafirma el alcance y la viabilidad de la investigación.
3. La pregunta debe hacer referencia a una sola cuestión. Formular una interrogante como ¿Cuál es el nivel de apropiación tecnológica y qué herramientas digitales usan las y los estudiantes del Telebachillerato Las Trancas? Remite a dos variables (el nivel de apropiación tecnológica y las herramientas digitales) que bien pueden abordarse en dos preguntas por separado.

En la medida en que se definan adecuadamente las preguntas de investigación, se estará en posibilidades de plantear los objetivos correspondientes.

Los objetivos de la investigación

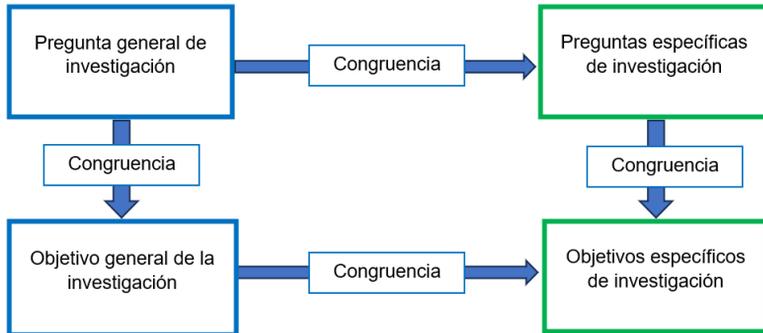
Cuando se ha delimitado el problema de la investigación y definido las preguntas que la orientarán, se está en posibilidades de construir los objetivos. En este proceso, debe cuidarse la congruencia interna. Debe existir una relación directa entre las preguntas y los objetivos planteados y, una vez que estos dos elementos han sido delimitados, también se cuidará la congruencia con las hipótesis y el diseño metodológico.

Los objetivos son enunciados que definen qué se pretende con la investigación. Deben atender a los

criterios de claridad (evitar ambigüedades), factibilidad (posibilidad de ser alcanzados, tanto en el tiempo como en relación con la metodología de la investigación) y pertinencia (abonar a una necesidad del saber en un área de conocimiento). Definen los alcances y límites del proceso investigativo, señalando los resultados que se esperan y los conceptos que se estudiarán. Al igual que en el caso de las preguntas de investigación, en los objetivos es posible distinguir entre general y particulares. Ojeda Ramírez et al. (2011) mencionan que el objetivo general explica globalmente y en una conceptualización amplia cuál es la pretensión del estudio. Por su parte, los objetivos particulares desglosan detalladamente los aspectos que se estudiarán; deben derivarse del objetivo general de la investigación y apoyar su logro.

La redacción del objetivo general puede orientarse por las siguientes preguntas: ¿qué se espera lograr con la investigación?, ¿qué se estudiará?, ¿con quién, ¿quiénes?, ¿dónde?, ¿cuándo?, elementos que guardarán congruencia con la pregunta general de investigación previamente establecidas. Mejía Henao (2019) agrega un elemento más: el para qué de la investigación. En cuanto a los objetivos específicos, éstos abordarán cuestiones relacionadas con las preguntas específicas. La Figura 4 ilustra esta relación.

Figura 4. Relación de congruencia entre los objetivos de la investigación.



Nota: Elaboración propia.

La enunciación de los objetivos de investigación inicia con un verbo en infinitivo. Arias (2012) propone una clasificación de verbos para redactar objetivos de investigación (Tabla 4) distinguiendo el nivel del estudio: exploratorio, descriptivo o explicativo, aunque hay que tener en cuenta que un objetivo es un fin; en este sentido hay muchos verbos que indican más bien acciones; es decir, estrategias, no fines, sino medios. Por ejemplo: analizar y calcular no son fines; habría que preguntarse para qué se analiza y para qué se calcula: ése sería el objetivo, que regularmente es para conocer, para saber, para determinar. Hay que tener mucho cuidado en esta parte.

Tabla 4. Verbos para redactar objetivos según el tipo de estudio.

Exploratorio	Descriptivo	Explicativo
Conocer	Analizar	Comprobar
Definir	Calcular	Demostrar
Descubrir	Caracterizar	Determinar
Detectar	Clasificar	Establecer
Estudiar	Comparar	Evaluar
Explorar	Cuantificar	Explicar
Indagar	Describir	Inferir
Sondear	Diagnosticar	Relacionar
	Examinar	Verificar
	Identificar	
	Medir	

Nota: Arias (2012, p. 44).

Por su parte, Martínez Rizo (2019) señala que la elaboración del objetivo general depende del tipo de diseño del estudio. Como ejemplo, en un estudio descriptivo sugiere distinguir el tipo de descripción que se pretende realizar. Si es de tipo simple, un verbo adecuado es “identificar”; si es correlacional, el objetivo puede enunciarse como “Analizar la relación entre (variables)”; finalmente, si el estudio es comparativo, el verbo recomendado es “comparar”. Para la elaboración de objetivos también se sugiere la revisión de la taxonomía de Bloom (Con sus actualizaciones, puede ser consultada en: <https://eduteka.icesi.edu.co/pdfdir/TaxonomiaBloomCuadro.pdf>), desarrollada por este autor en 1956. Bloom clasifica los verbos a partir del nivel de complejidad:

- Menor complejidad: Nivel 1. Conocimiento. Ejemplos de verbos: reconocer, saber, identificar, exponer. Nivel

...en el proceso de investigación

2. Comprensión. Ejemplos de verbos: Interpretar, clasificar, explicar, describir, revisar.
- Mayor complejidad: Nivel 3. Aplicación. Ejemplos de verbos: demostrar, aplicar, examinar, asociar. Nivel 4. Análisis. Ejemplos de verbos: comparar, analizar, contrastar, distinguir, relacionar, discriminar. Nivel 5. Síntesis. Ejemplos de verbos: Desarrollar, formular, integrar, diseñar, planificar, crear. Nivel 6. Evaluación. Ejemplos de verbos: establecer, medir, valorar, probar.

La taxonomía puede ser aplicable para objetivos de aprendizaje y para objetivos de investigación. En el caso de los posgrados se espera que los niveles de complejidad sean los más altos; particularmente en el caso del doctorado. En este nivel se esperan objetivos de investigación que favorezcan la generación de aportaciones valiosas al campo de conocimiento en que se ubican, a partir de procesos de síntesis y evaluación de los objetos de estudio.

Así, la redacción de los objetivos considera cuestiones propias de la delimitación del problema, pero también en relación con el tipo de estudio, alcance y diseño metodológico de la investigación. Cabe señalar que como ocurre con las preguntas de investigación —y en congruencia con ellas— los objetivos pueden modificarse o ajustarse en el desarrollo de la investigación. Como señala Rojas Soriano (2010) la investigación implica un proceso dialéctico, en donde se vuelve constantemente sobre los planteamientos, de lo teórico a lo empírico.

Para cerrar con este apartado, se presentan enseguida los errores más comunes en la formulación de los objetivos de investigación, como una manera de alertar a quien se encuentra en este proceso y estar en posibilidades de evitar caer en ellos:

1. Plantear un objetivo demasiado amplio. Ejemplo: “Conocer los efectos de la baja autoestima en el aprendizaje”. En este objetivo no se clarifica la población de estudio, ni el contexto específico de la investigación.
2. Formular objetivos con pocas o nulas posibilidades de ser alcanzados. Ejemplo de este error puede ser, en una tesis de licenciatura, tener como objetivo de investigación: “Construir una propuesta educativa enfocada en los trastornos emocionales del alumnado”. En este caso, por tratarse de un trabajo de nivel profesional, el objetivo es demasiado ambicioso y difícil de lograr, considerando el tiempo que los programas de pregrado establecen para el desarrollo de investigaciones con fines de titulación (regularmente de un año).
3. Incluir dos verbos en un mismo objetivo. Ejemplo: “Conocer y analizar las diferencias entre una escuela rural y una urbana”. Este objetivo, además de ser demasiado amplio (caso ilustrado en el primer inciso), contiene dos verbos: Conocer y analizar. La recomendación es definir dos objetivos (uno por cada verbo) si la investigación así lo requiere, considerando que seguramente uno de ellos puede dar lugar a un objetivo general y el otro a uno específico.

Conclusiones

Los elementos desarrollados en este capítulo (delimitación del problema, formulación de las preguntas y enunciación de objetivos) son procesos fundamentales para el desarrollo de un trabajo de investigación académica científica, ya que de ellos emanarán las bases para el marco conceptual teórico y la metodología

...en el proceso de investigación

del estudio. Estos tres elementos se recuperarán además cuando quien investiga presente el análisis de resultados y las conclusiones, pues en ellas se tiene que recuperar de manera general las respuestas a las preguntas y el alcance de los objetivos.

En el desarrollo del proceso investigativo es posible que las preguntas y los objetivos sean modificados, a partir de la revisión teórica, los hallazgos iniciales las necesidades de replantear la delimitación del problema. Todo ello implica un proceso dialéctico, en el cual la recursividad y la habilidad creativa de quien realiza el estudio son necesarias, en aras de la construcción de una investigación pertinente y factible.

Referencias

Fidias Arias, *El proyecto de investigación. Introducción a la metodología científica*, Venezuela, Editorial Episteme, 2012.

http://www.formaciondocente.com.mx/06_RinconInvestigacion/01_Documentos/El%20Proyecto%20de%20Investigacion.pdf

Juan Camilo Arias Castrillón, “Plantear y formular un problema de investigación: un ejercicio de razonamiento” *Revista Lasallista de investigación*, núm. 1, julio 2020, pp. 301-313.

<http://www.scielo.org.co/pdf/rlsi/v17n1/1794-4449-rlsi-17-01-301.pdf>

Eduardo Enrique Espinoza Freire, “El problema de investigación”, *Revista Conrado*, núm. 64, julio-septiembre 2018, pp. 22-32.

<https://conrado.ucf.edu.cu/index.php/conrado/articulo/view/808>

- Felipe Martínez Rizo, *El nuevo oficio del investigador educativo. Una introducción metodológica*, México, Universidad de Aguascalientes, 2019.
- Mejía Henao, J. D. (2019, 17 de agosto). *Cómo hacer objetivos generales y específicos de investigación* [Video]. YouTube.
https://www.youtube.com/watch?v=X_xb7JEXBtI
- Mario Miguel Ojeda Ramírez, José Enrique Díaz Camacho, Clara Apodaca Victoria y Israel Trujillo Landa, *Metodología del diseño estadístico*. México, Universidad Veracruzana, 2011.
<http://libros.uv.mx/index.php/UV/catalog/view/TU162/734/898-1>
- Cecilia María Patiño y Juliana Carvalho Ferreira, “Developing research questions that make a difference”, *Jornal Brasileiro Pneumologia*, núm. 6, noviembre-diciembre 2016, pp. 403.
<http://dx.doi.org/10.1590/S1806-37562016000000354>
- Carlos Alberto Ramos Galarza, “La pregunta de investigación”, *Avances en Psicología*, núm. 1, enero-julio 2016, pp. 23-31.
<https://revistas.unife.edu.pe/index.php/avancesenpsicologia/article/view/141>
- Raúl Rojas Soriano, *El proceso de la investigación científica*, México, Trillas, 2010.

Diseño estadístico: tipo de estudios, unidades, variables y escalas de medición

MARIO MIGUEL OJEDA RAMÍREZ
LUIS GERARDO DOMÍNGUEZ REYES

Introducción

Quien investiga debe hacer una serie de consideraciones al iniciar el diseño del estudio o investigación, las cuales serán claves en la determinación de la estrategia metodológica. El marco referencial que tiene ante sí es lo que se conoce como metodología de investigación. Hay ciertamente muchos libros y manuales que consideran estos aspectos, pero con frecuencia, sobre todo al iniciar esta carrera, se encuentra ante un amplio abanico de posibilidades. En este capítulo nos hemos planteado el reto de presentar una revisión sucinta que considere los principios y el marco referencial que plantea la estadística para el diseño de una investigación. Se hace una revisión de los tipos de estudios, la clasificación general de los objetivos de la investigación, así como lo relacionado con el proceso de medición. No sin antes enfatizar en que una nítida comprensión y delimitación del problema, deja con mayor claridad cómo debe ser el diseño de la investigación. Lo que aquí se ha considerado como diseño, que se plantea en un protocolo, es el mapa que permitirá ir del punto inicial al punto final.

Tipos de estudio

Ander-Egg (1972) distingue la investigación básica de la aplicada. La primera se orienta a desarrollar o profundizar conocimientos de algún fenómeno, sin procurar las posibles utilidades. La segunda, utiliza los conocimientos científicos para un propósito ulterior. Martínez Rizo (2019) organiza la investigación básica en cuatro categorías: viva, documental, observacional y no observacional. La investigación viva se refiere al conjunto de investigaciones que generan información nueva, basada en datos no recogidos anteriormente. Por el contrario, la investigación documental responde a sus objetivos mediante el uso de datos obtenidos previamente. Los estudios observacionales son aquellos en los que no hay manipulación de variables o factores en la obtención de las unidades de estudio. Los estudios no observacionales no se limitan a la observación del fenómeno en estudio, sino que manipulan por lo menos una variable en la obtención de las mediciones a fin de observar su efecto en otra variable.

Batthyány y Cabrera (2011) agrupan las investigaciones en cuatro diseños según sus objetivos: exploratorio, descriptivo, predictivo y evaluativo. Las investigaciones exploratorias se caracterizan por buscar examinar un objeto de conocimiento poco estudiado o no abordado antes. Las investigaciones descriptivas pretenden caracterizar a las unidades de análisis. Los diseños explicativos se orientan a identificar las causas de los eventos físicos o sociales. Se busca amplificar la explicación del por qué ocurre un fenómeno o por qué dos o más variables están relacionadas. Por otra parte,

los diseños predictivos, el comportamiento futuro de dos o más variables.

Ojeda Ramírez et al. (2011) describen tres tipos de estudios estadísticos: observacionales, experimentales y de muestreo. Los estudios observacionales son aquellos cuyas unidades de estudio están dadas en la investigación, es decir las variables no son manipuladas. Sólo es posible observarlas. En los estudios experimentales las unidades de estudio se agrupan mediante un mecanismo aleatorio y se asigna un tratamiento para cada grupo. Los estudios muestrales se caracterizan por seleccionar un subconjunto de la población para observar las variables de investigación. Así mismo, proporcionan una serie de recomendaciones fundamentales para diseñar un estudio estadístico, antes de apresurarse a conseguir o a obtener datos. En primer lugar, se destaca que para aplicar o apropiarse de un diseño estadístico para su investigación, es imprescindible que quien investiga defina con precisión: su problema y utilidad, cuál es su colectivo de estudio o población, que conjunto de variables interesa medir, y cuál será la naturaleza del estudio (meramente descriptiva o analítica). Del mismo modo, recomiendan pensar en cuáles serán los resultados que se obtendrán y cómo se interpretarán. Todos estos elementos sirven como una brújula para comenzar a considerar si se tiene claro qué es lo que se va a estudiar, identificar la presencia de posibles factores u otras variables que puedan impactar sobre el análisis, y corregir por si no se han tomado en cuenta; así mismo, con estos elementos se podrá definir, en el caso de los estudios de muestreo u observacionales, el tipo de diseño que se utilizará y qué tamaño de muestra se requiere.

Ynoub (2011) distingue tres tipos de investigación: descriptivas, explicativas e interpretativas. Las investigaciones descriptivas tienen como objetivo identificar: (a) el comportamiento de una o varias variables; (b) el comportamiento conjunto de variables al modo de correlaciones que no impliquen causalidad, entendida como variación conjunta entre dos acontecimientos; o (c) la combinación de variables entre las variables.

Las investigaciones explicativas buscan determinar “si el comportamiento de una variable (o varias) puede funcionar como factor o causa explicativa del comportamiento de otra” (Ynoub, 2011, p. 85). Esta relación se puede probar mediante dos tipos de estudios: (a) correlacionales que implican (o pueden implicar) causalidad, y (b) experimentales. Una correlación ocurre cuando los valores de una variable varían sistemáticamente de manera directa o inversa respecto a los de otra. Una correlación implica causalidad cuando la presencia de una variable es anterior en el tiempo a otra, y el conocimiento del mundo autoriza para poder postular esa relación causal. Si no se cumplen estos criterios la correlación no implica causalidad, lo que la ubicaría dentro del grupo de las investigaciones descriptivas. Las investigaciones experimentales son aquellas que (a) prueban hipótesis que formulan relaciones de tipo causal entre dos o más variables; y (b) se realizan bajo condiciones creadas y controladas por quien investiga.

Las investigaciones interpretativas se orientan a la interpretación o comprensión de los fenómenos. Ynoub (2011) las agrupa en dos categorías: cualitativas y de fenómenos culturales. La investigación cualitativa es desarrollada fundamentalmente en las ciencias sociales

...en el proceso de investigación

y humanas. Se nutre de orientaciones filosóficas interesadas en la comprensión de fenómenos históricos humanos y subjetivos. Estos estudios se caracterizan por su comprensión de los fenómenos. Consisten en brindar información procurando integrar la perspectiva de los sujetos o los fenómenos estudiados en una situación determinada, atendiendo la mayor parte de aspectos que puedan resultar relevantes. Además, las investigaciones cualitativas, observan los datos y evidencias en sus contextos naturales, es decir, en la situación en la que se presentan habitualmente; están asociadas a entrevistas, uso de bitácoras e observación y otros métodos antropológicos.

Unidad de análisis, unidades de estudio y colectivo de estudio

Son un elemento estructural de la hipótesis que alude al objeto de estudio. Estas tienen atributos o características distintas total o parcialmente al de otras unidades (Ortiz Uribe, 2003). Son entidades identificables en algún tiempo y o espacio; por tanto, son numerables, es decir, se puede identificar el número de unidades con el que efectivamente se trabajó (Ynoub, 2011). Por ejemplo, en un estudio de egresados realizado en México por Ojeda Ramírez et al. (2014), las unidades de análisis fueron quienes han egresado del posgrado de la Universidad Veracruzana.

Las unidades de estudio son las unidades básicas sobre las que se registran los datos; es decir, son las personas o entidades que se miden. Por ejemplo, en la investigación sobre los egresados serían las personas que cumplen con la definición de egresado; es decir,

aquellos que han completado un programa de estudios de este nivel. A las unidades de estudio a veces también se les llama unidades experimentales, si es que el estudio es de tipo experimental.

Finalmente, el colectivo de estudio es generalmente la muestra o el conjunto de unidades de estudio; cuando los estudios son de muestreo se tiene por un lado la población de muestreo, que es el conjunto de unidades del que se extrae la muestra. En los estudios analíticos se le denomina población de referencia o población foco (*target population*).

Dentro de la investigación, la estadística contribuye a diseñar instrumentos de captación como en las encuestas, los cuestionarios, o los test, entre otros. Así como en la implementación de técnicas y métodos para conseguir muestras que sean representativas. Llevar a cabo de forma correcta la recolección de los datos constituye un aspecto fundamental; si los datos obtenidos no son de calidad, muy probablemente los resultados finales tampoco lo serán. Sin embargo, quien investiga no siempre tendrá la oportunidad de participar en la recolección de los datos. En estos casos, no deberá pasar por alto aspectos importantes como la fuente de los datos, cómo se llevó a cabo su registro o el instrumento que se utilizó. Todo esto con la finalidad de identificar las fuentes de variación y factores no obvios que puedan influir sobre la medición, los datos y, por ende, en las fases subsecuentes en la investigación.

Es común que las unidades de estudio se seleccionen en forma caprichosa en las investigaciones, atendiendo a la participación voluntaria o por cuotas, como sucede en las encuestas de opinión, las cuales tienen baja credibilidad, al tratarse de “muestras circunstanciales o

erráticas” (Calderón García et al., 2019). Por esta razón la metodología estadística proporciona los principios y técnicas de muestreo. Se debe tener suficientemente claro cuál es la población objetivo, a partir de este conocimiento se escoge un método para seleccionar una muestra que pueda ser considerada como estadísticamente representativa. Cabe destacar que el muestreo puede ser probabilístico o no probabilístico; y en ambos casos se pueden garantizar muestras representativas y de buena calidad.

De acuerdo con las características del problema y el diseño estadístico, convendrá utilizar un muestreo u otro. Los más comunes son el muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo por conglomerados y el muestreo sistemático; todos ellos entran dentro de los muestreos probabilísticos. Por el lado de los “no probabilísticos”, algunos de los más conocidos son el muestreo por juicio, por conveniencia, el muestreo bola de nieve, entre otros. Estos últimos, aunque en ciertos contextos sean más convenientes, son susceptibles de ser cuestionados y en ocasiones proporcionar datos sesgados. Vale la pena hacer hincapié en que la selección del tipo de muestreo está estrechamente relacionada con qué tan bien se conozca a su colectivo de estudio, así como a las variables que son de interés. Al conocer la estructura de la población objetivo, será posible identificar si está dividida por estratos, o si sus unidades se encuentran agrupadas en conglomerados, o si simplemente conviene hacer un muestreo de la forma más simple.

Se ha popularizado el muestreo aleatorio estratificado, en el cual se seleccionan muestras dentro de cada estrato, para posteriormente combinar esta información

con el fin de hacer inferencias acerca de toda la población, pero así mismo para comparar los resultados que se obtienen por estrato; de igual forma, a veces se segmenta la población, definiendo lo que se llama dominio de estudio. Por ejemplo, pueden visualizarse las diferencias de opinión entre hombres y mujeres, o entre jóvenes y adultos mayores. Existen diferentes diseños de muestra los cuáles pueden adaptarse a distintas necesidades de cada investigación. Así como la muestra por conglomerados se ajusta mejor cuando los sujetos de estudio pertenecen a grupos pequeños, el muestreo por cuotas cuando se trata de abarcar un determinado número de personas con ciertas características y el muestreo sistemático cuando se antepone criterios básicos de orden o alineamiento a partir de los cuales se seleccionan los participantes de la muestra.

Tipos de variables

Todo conocimiento es imperfecto y analítico. Es imposible conocer la totalidad de la realidad en un sentido estrictamente holístico (Martínez Rizo, 2019). Así que sólo es posible conocer ciertos aspectos de la realidad, que hay que abstraer para poderlos captar. Toda investigación inicia precisando los aspectos de la realidad que le interesa estudiar; éstos son llamados variables. Por lo tanto, una de las habilidades esenciales de quien investiga es la identificación de las variables.

Una variable es un “símbolo al que se le asignan valores”, “alguno de los rasgos o de las propiedades que ayudan a distinguir la ocurrencia de algo”, “cualquier característica cuyo valor puede cambiar de un objeto a otro en la población” (Kerlinger, p. 36). Estas definiciones

muestran que la característica esencial de la variable es la diferencia de valor que toma de una unidad de estudio a otra. Es importante distinguir entre cambio y diferencia. Todo cambio implica una diferencia, pero no toda diferencia un cambio. El cambio es el resultado de transitar de una situación o característica a otra. Se puede decir que un objeto de conocimiento ha cambiado porque su estado, en algún aspecto, es diferente al de antes. En contraste, la diferencia alude a una propiedad que se distingue de otra. Ejemplo: A es blanco y B es negro. Los dos objetos tienen propiedades diferentes, pero no necesariamente por ser el resultado de un cambio. Ambos pudieron tener esos colores desde sus inicios.

Por lo tanto, la variable es un aspecto cuya propiedad puede diferir de un elemento a otro. Bunge (2005) define propiedad como “rasgo o característica que posee un objeto conceptual o material” (p. 174). El aspecto sexo puede tener la característica hombre o mujer, o el atributo número de empleos que ha tenido una persona puede asumir valores como uno, dos, etc. Ejemplos de variables son: estado conyugal, nivel educativo, número de hijos o ingresos. Una propiedad por sí misma no constituye una variable. El conjunto observado debe admitir por lo menos dos valores distintos. Cuando el atributo o característica no varía se le llama constante (Ortiz Uribe, 2003). Así, por ejemplo, la edad es una variable sólo si en el conjunto observado hay por lo menos dos edades distintas. Si en la población, muestra o colectivo de estudio todos los elementos tienen la misma edad, entonces no es una variable sino una constante.

Existen varias formas de clasificar a las variables. Una de las más comunes es según el tipo de valores que expresa. De acuerdo con esta clasificación las variables pueden ser cualitativas, cuantitativas, dentro de éstas, discretas y continuas; las cualitativas y discretas también pueden ser denominadas categóricas. Las variables cualitativas son aquellas cuyos valores son cualidades o categorías no numéricas (Lind et al., 2015). Por ejemplo, la variable sexo es cualitativa puesto que los valores que puede asumir (femenino o masculino) no son numéricos. Por lo general, en las variables categóricas se hace el conteo de las observaciones por categoría y se determina el porcentaje de cada una. Las variables cuantitativas son aquellas que expresan una cantidad numérica (Mendehall et al., 2015). Por ejemplo, el número de cursos MOOC que ofrecen las universidades mexicanas o la matrícula de hablantes de lenguas indígenas en universidades interculturales.

Las variables cuantitativas pueden ser discretas o continuas (Lind et al., 2015). Las variables discretas abarcan un número finito o contable de valores, por lo que existe brecha entre los valores. Por ejemplo, el número de hijos. Una persona puede tener 3 o 4 hijos, pero no 3.14 hijos. Se dice que hay una brecha entre los valores posibles de la variable. En contraste las variables continuas asumen “un número finito o contable de valores” (Mendehall et al., 2015); se les llama continuas porque este conjunto finito se debe a la graduación o precisión de la escala de este conjunto de valores, los cuales suceden en un intervalo; es decir, en un continuo de valores.

Otra clasificación de variable se organiza en función de la forma en que se relacionan las variables entre sí. A

menudo esta clasificación se utiliza en términos de causalidad como lo plantea Kerlinger (2002) o Dieterich (2011). En los diseños experimentales es posible atribuir causalidad debido al control de las variables y los grupos, no así en otro tipo de estudios. En ocasiones las relaciones entre las variables son de influencia o contribución, por lo que aquí se utilizarán en estos términos, sin que implique necesariamente causalidad. A partir de estas consideraciones se puede definir la variable dependiente (o explicada) como aquella cuyos valores son modificados o influidos por otra variable, acorde a lo planteado por quien investiga. Las variables independientes (o explicativas) son aquellas que modifican o contribuyen en los valores de otra variable. Esta distinción permite comunicar su planteamiento sobre la forma en que se relacionan las variables en el marco de su estudio.

Es importante advertir que para clasificar las variables como dependientes o independientes no existen criterios intrínsecos (Ynoub, 2011). Tampoco se requiere que la relación entre estas variables esté probada. Basta con que en la hipótesis de investigación se plantee cuál variable se pretende explicar y cuál(es) variable(s) contribuye(n) en su comportamiento. Así, por ejemplo, si en un estudio se supone que el sexo influye en las remuneraciones que percibe un trabajador, la variable remuneración es dependiente, puesto que a partir de marco referencial (conceptual, teórico, revisión de la literatura) se considera que está influida por el sexo. Por su parte, el sexo es considerado como variable independiente, puesto que se plantea que sus valores (masculino y femenino) modifican o contribuyen en la remuneración de trabajador.

Escalas de medición

En un sentido amplio la medición consiste en asignar números o categorías a propiedades de objetos de acuerdo con reglas (Kerlinger, 2002). Estas reglas son muy diversas. Stevens (como se citó en Mendehall et al., 2017) distingue cuatro tipos de escalas según las propiedades del sistema numérico, lo que se aprovecha para la regla que se usa en la asignación: nominal, ordinal, de intervalo y de razón.

La escala nominal emplea números como nombres para las clases de objetos (Pérez-Tejada, 2009). Así, por ejemplo, se podría usar 1, para sexo masculino y 2 para sexo femenino (o viceversa). Esta escala no permite establecer un tipo orden entre las categorías de las variables (Batthyány y Cabrera, 2011).

En la escala ordinal, los números son asignados con la propiedad ordinal del sistema numérico (Mendehall et al., 2017). Los valores implican un orden; es decir, muestran la posición de un elemento en una sucesión finita. Por lo tanto, se asignan valores numéricos a una propiedad del objeto de modo que reflejen niveles crecientes de esa propiedad. Por ejemplo, en un estudio de egresados se puede asignar números a la satisfacción de los ocupados con su empleo de la siguiente forma: 1 a Nada, 2 a Poco, 3 a Suficiente, 4 a Mucho y 5 a Totalmente. No se puede afirmar que la distancia entre el que responde 1 y 2 es la misma distancia entre 4 y 5, pero sí permite reconocer que el valor 5 es mayor que el 4 en esta variable.

Las escalas de intervalo no sólo muestran orden, sino que también establecen que la distancia entre los valores. Un ejemplo es la temperatura en grados Celsius. La distancia entre 15 °C y 20 °C es igual que entre 30 °C y 35 °C. En estas escalas el cero es arbitrario y no refleja la ausencia de la propiedad que se mide. Es decir 0 °C no refleja la ausencia de calor, o una talla cero en una ropa no indica cero centímetros (o pulgadas) de longitud.

Las escalas de razón tienen todas las propiedades de las escalas de intervalos, pero además el cero sí refleja la ausencia de una cualidad. Ejemplos son la altura, peso, duración, etc.; en general se refieren a todas las variables que se miden con instrumentos de precisión con cero absoluto.

En resumen, los tipos de escalas se distinguen entre sí por las propiedades numéricas que admiten en la medición (Mendenhall et al., 2017). La escala nominal utiliza la propiedad de identidad de números, por lo que es útil para indicar la pertenencia a un grupo (o clase) o clasificación. Las escalas ordinales, utilizan la propiedad ordinal; es decir, el hecho de seguir una secuencia. Las escalas de intervalo, además de las propiedades de identidad de números y ordinal, tienen la del establecimiento de la distancia. Las escalas de razón tienen todas las propiedades numéricas, las tres propiedades anteriores y la de referencia al cero absoluto.

Distinguir las escalas de medición es útil en las investigaciones puesto que el uso de la estadística se ve limitado por el tipo de escalas que utiliza. Mendenhall et al. (2017) afirma que las escalas de intervalo razón permiten el uso de los procedimientos más robustos llamados paramétricos. En cambio, las escalas

nominales y ordinales utilizan regularmente los procedimientos no paramétricos. En consecuencia, el tipo de escala de medición determina, en gran medida, el tipo de métodos estadísticos que se pueden emplear.

A manera de conclusión

El proceso de investigación tiene su punto nodal en el diseño del estudio; esto es, la determinación del tipo de estudio y la determinación del proceso de medición. ¿Cómo se seleccionarán las unidades de estudio? ¿Cómo se obtendrá su información? ¿Qué unidades se medirán?, ¿qué variables? y ¿en qué escalas?, son algunas de las preguntas que la persona que realiza el proceso debe contestar. La ciencia tiene en la investigación factual un poderoso instrumento para aumentar el conocimiento existente. En este marco la metodología estadística, con los principios y procedimientos que determinan el diseño del estudio es fundamental. Por tal motivo las investigadoras y los investigadores deben implicarse en este tema y no pasar por alto que estos aspectos son fundamentales y constituyen la médula espinal de la investigación. Si el proceso para elaborar el diseño estadístico del problema se hace con poca cautela, es probable que los datos no sean los adecuados para el propósito de la investigación, que contengan más errores de los que habitualmente se esperan y detalles en su estructura que comprometan y obliguen a quien investiga a cambiar los objetivos iniciales y replantear el trabajo. Por consiguiente, este proceso debe planificarse escrupulosamente. La recolección de los datos es un procedimiento que, por cuestiones de tiempo y dinero, casi nunca es factible

...en el proceso de investigación

realizarlo una y otra vez. Se debe buscar que este proceso pueda aplicarse de una forma práctica, sencilla y económica, buscando maximizar los recursos de los que se dispone. En el entendido de que ya se ha diseñado un estudio y se han recopilado los datos del colectivo de estudio, proseguiría inspeccionar la estructura de los datos y elaborar el proceso de depuración o curación de los datos.

Referencias

- Ezequiel Ander-Egg, *Introducción a las técnicas de investigación social para trabajadores sociales*, Humanitas, 1972.
- Karina Batthyány y Mariana Cabrera, *Metodología de las ciencias sociales. Apuntes para un curso inicial*, Uruguay, Universidad de la República, 2011.
- Mario Bunge, *La investigación científica: su estrategia y su filosofía*, Argentina, Siglo veintiuno editores, 2005.
- Arturo Calderón García, Cristian Bayes Rodríguez y Luis Valdivieso Serrano, *Guía de investigación en ciencias y en ingeniería*, Perú, Pontificia Universidad Católica de Perú (PUCP), 2019.
<https://cdn02.pucp.education/investigacion/2016/06/29210737/Guia-de-Investigacion-en-Estadistica.pdf>
- Hienz Dieterich, *Nueva guía para la investigación científica*, México, Grupo Editor Orfila Valentini, 2011
- Fred Kerlinger, *Investigación del comportamiento: Métodos de investigación en ciencias sociales*, Chile, McGraw Hill, 2002.

- Douglas Lind, William Marchal y Samuel Wathen, *Estadística aplicada a los negocios y la economía*, México, McGraw-Hill, 2015.
- Felipe Martínez Rizo, *El nuevo oficio del investigador educativo. Una introducción metodológica*, México, Universidad Autónoma de Aguascalientes, 2019.
- William Mendehall, Robert Beaver y Barbara Beaver, *Introducción a la probabilidad y estadística*. México, Cengage Learning Editores, 2015.
- William Mendehall, Robert Beaver y Barbara Beaver, *Probabilidad y estadística para las ciencias sociales, del comportamiento y la salud*, México, Cengage Learning Editores, 2017.
- Mario Miguel Ojeda Ramírez, José Enrique Díaz Camacho, Clara Apodaca Victoria y Isarel Trujillo Landa, *Metodología del diseño estadístico*. México, Universidad Veracruzana, 2011.
- Frida Ortiz Uribe, *Diccionario de metodología de la investigación científica*, México, Limusa, 2003.
- Haroldo Elorza Pérez-Tejada, *Estadística para las ciencias sociales, del comportamiento y de la salud*. México, Cengage Learning Editores, 2008.
- Roxana Cecilia Ynoub, *El proyecto y la metodología de la investigación*, Argentina, Cengage Learning, 2011.

Estructura y tratamiento preliminar de los datos

ARIEL ANTONIO LÓPEZ SALAS
CARLOS GARCÍA TRUJILLO

Introducción

La investigación puede entenderse como un oficio intelectual. Con frecuencia requiere de la aplicación de la estadística, para que, en sintonía con la teoría, se coadyuve al avance del conocimiento. Para llevar a cabo el análisis estadístico es necesario haber recopilado datos, producto de los procesos de medición planteados anteriormente; estos datos pueden provenir de encuestas, experimentos, así como de sistemas de información específicos para captarlos. Cualquier conjunto de datos puede contener irregularidades, daños o inconsistencias que pueden comprometer la confiabilidad de los resultados; es decir, su veracidad. Inspeccionar los datos antes de procesarlos permite pensar en su fiabilidad, sus fortalezas y debilidades; es decir, en su calidad.

Contar con datos de buena calidad posibilita que los hallazgos sean susceptibles de traducirse en conocimiento relevante. Antes de comenzar con el análisis estadístico, los datos requieren ser inspeccionados, procesados tentativamente y depurados con suficiente cautela, evitando las decisiones arbitrarias. Esto permitirá pasar de una matriz con datos “crudos”

(como comúnmente se denomina) a una base de datos limpia. Con ella se podrán implementar los procedimientos de análisis para cumplir satisfactoriamente con los propósitos que se ha planteado quien investiga. Con el mismo nivel de importancia, se debe comprender cómo está estructurada la base de datos; es decir, cómo son las variables, las relaciones que guardan entre sí, y el rol que van a desempeñar cada una de ellas dentro del análisis.

En el proceso de limpieza es probable que quien investiga se encuentre con datos faltantes, atípicos, errores de medición o de captura; también se pueden encontrar inconsistencias de diferente tipo. Por ello, antes de emprender camino en el análisis, se deberá estar familiarizado con las múltiples situaciones que pueden presentarse al explorar los datos crudos y, sobre todo, cómo pueden resolverse. Todo ello engloba la fase del pre-procesamiento o preparación de los datos. Como mencionan Zhang et al. (2010), esta etapa abarca a todas aquellas técnicas que, en general, mejorarán la calidad de los datos. De tal manera que al aplicar las técnicas estadísticas puedan reflejar la información que más se asemeja a la realidad investigada.

Hoy en día, dentro de las investigaciones fácticas, la construcción y generación del conocimiento debe estar sustentada y fundamentada desde la argumentación del buen uso de la metodología estadística. El campo de la investigación ha cambiado en las últimas décadas. Destaca la facilidad para acceder a nuevos programas computacionales, así como la aplicación y surgimiento de nuevas técnicas estadísticas para analizar evidencias; es decir, datos cuantitativos y cualitativos. Como bien lo

describen Lloret-Segura et al. (2014), quien investiga puede verse desconcertado ante la variedad de alternativas que han aparecido. Debe entonces abrirse camino tomando decisiones en cada paso; en ocasiones desde la experiencia, trabajos similares, o en una combinación de ambas. Por consiguiente, debe apropiarse de las herramientas y de los conocimientos para el tratamiento de los datos, así como de sus consideraciones, ventajas y desventajas involucradas en la aplicación. En este capítulo, se abordan algunas de las técnicas, métodos y procedimientos para llevar a cabo la “preparación de los datos”. Dicha preparación se desglosa en la comprensión de la estructura, la depuración y limpieza de los datos, y la exploración inicial, también llamado análisis exploratorio. Se hace una explicación detallada sin caer en los tecnicismos.

Comprensión de la estructura de los datos

Al hablar de archivos de datos (lo que en un lenguaje coloquial se conoce como bases de datos) se hace referencia a características o atributos específicos de un conjunto de individuos u objetos, los que se constituyen en el colectivo de estudio. El software estadístico, generalmente, considera esta base como una matriz de datos en la que cada columna representa una variable (característica) y cada fila un individuo, objeto o entidad del colectivo. Cerciorarse de que los datos están representados en este formato, es el primer paso. Se debe estar consciente de la importancia de prestar atención hacia la estructura de la base de datos, antes de comenzar con cualquier otro procedimiento. Como describe Ojeda (1988), el entendimiento claro de su

estructura permite manejar los datos con mayor eficiencia, así como un mejor aprovechamiento estos. De igual manera, funciona para dar dirección a la investigación y en muchas ocasiones hace que la elección de la técnica o método estadístico a emplear sea más evidente.

Las bases de datos pueden ir de estructuras simples hasta algunas de mayor complejidad. Sin embargo, como se verá a continuación, las estructuras complejas son las que aparecen comúnmente en las investigaciones. La naturaleza de éstas está determinada únicamente por dos tipos de variables: las variables de agrupación y las variables de medición. Las primeras se caracterizan por distinguir o identificar a diferentes tipos de individuos u objetos de estudio (sexo, edad, área académica, región, etc.). Las variables de medición, generalmente, son las características que interesan en el estudio (calificaciones, aptitudes, número de horas dedicadas al estudio, hábitos de lectura, etc.). Es importante aclarar que comúnmente las variables de agrupación son categóricas, en cambio las variables de medición pueden ser tanto cualitativas como cuantitativas.

La estructura más simple es la que se contiene solamente un grupo y una variable de medición. Es decir, la variable de agrupación no está presente. Por tanto, no hay manera de distinguir entre grupos de individuos. Esto podría ejemplificarse como el caso en el que se tienen las calificaciones del examen de ingreso a la universidad de 1000 estudiantes; únicamente se tiene un grupo (que son los estudiantes) y un atributo o característica (las calificaciones). A esta estructura se le pueden ir añadiendo niveles que generen una base de datos más compleja e interesante. Para ejemplificarlo, se podría

tomar el mismo escenario con la diferencia de que ahora se distinguen dos grupos de estudiantes, clasificados por el sexo. Así, se tiene una variable de agrupación y una variable de medición. La primera de ellas permite distinguir entre dos grupos de estudiantes y la segunda una medición. De forma inmediata se puede hacer una generalización: en lugar de tener únicamente dos grupos, se podrían tener más; en general se puede hablar de k grupos de estudiantes, que podrían representar el área académica a la cual presentaron su examen, y la misma variable de medición. Este tipo de configuraciones se pueden incluir dentro de las llamadas “estructuras simples”.

En lugar de tener únicamente una variable de agrupación, se podrían tener dos o incluso más; lo mismo ocurre para el caso de las variables de medición. Es en este punto donde ya se está hablando de las estructuras consideradas como “complejas”. Por ejemplo, ahora como variables de agrupación se considera el sexo y el área académica del estudiante, y como variables de medición su calificación y el tiempo en minutos que tardó para terminar el examen. A este hipotético escenario se le podrían seguir agregando más características, tanto para distinguir a los individuos como para registrar su respuesta. Estos dos grupos de variables guardan una relación que, con frecuencia, suelen denominarse variables independientes (las de agrupación) y dependientes las de medición.

A medida que esta base se vuelve más rica, se requiere de mayores competencias tanto intelectuales como estadísticas para guiar adecuadamente el análisis y aprovechar al máximo la información. También de un mayor esfuerzo para construir conclusiones que tomen

en cuenta el comportamiento de las variables en su conjunto, sus asociaciones e interrelaciones. Las estructuras de mayor complejidad exigen de un enfoque más nítido y preciso por parte de quien investiga, respecto a qué es a lo que se quiere llegar a conocer o saber a través de los análisis.

Las bases de datos con estructuras simples pueden ser analizadas a través de técnicas más directas, dado que las posibilidades de la base son reducidas. Cabe aclarar que esto no se traduce en que sean poco útiles. Cada estructura funcionará de acuerdo con los propósitos del estudio y de los métodos estadísticos que se pretenda utilizar. Por otra parte, las estructuras complejas colocan a quien investiga en un sendero con más opciones. Sin embargo, si se quiere llegar a resultados valiosos, se deberá trabajar escrupulosamente con los datos antes de transitar a los análisis definitivos.

Dentro de las estructuras complejas, es habitual que se tengan variables que no son realmente necesarias o que son poco informativas para el estudio. Además, al contar con un gran número de variables, se debe contar con un mayor número de casos, individuos o entidades, si se quiere que los análisis produzcan resultados relevantes. Todos estos aspectos deben ser cuidadosamente tratados antes de intentar cualquier análisis, sobre todo aquellos multivariantes o que pretenden ajustar e interpretar modelos estadísticos.

Una vez que se ha conocido la estructura de los datos, es momento de transformar el conjunto de datos de su estado primario a una base de datos que sea adecuada para realizar los análisis exploratorios y los definitivos. Esto implica una depuración y una limpieza

...en el proceso de investigación

exhaustiva. Es decir, la eliminación de las inconsistencias, los valores ausentes y los errores de medición y, en ocasiones, modificar las escalas.

Tratamiento de datos faltantes

Uno de los problemas más comunes es que el conjunto de datos contenga datos faltantes. Los valores perdidos, generalmente, se le pueden atribuir a factores externos o ajenos a quien investiga. Ya sea porque las unidades de estudio son poco accesibles, los datos que se solicitan son difíciles de obtener, o atentan contra la privacidad de la información de los individuos, entre otras razones. Aunado a esto, hay otras causas sobre las cuáles se debe tener la capacidad para contrarrestar sus efectos: mala redacción en las preguntas, ítems desconectados de los objetivos de la investigación, y cuestionarios muy extensos, por mencionar las más frecuentes.

Dicho problema es determinante por una razón: muchas de las técnicas y métodos estadísticos son sensibles a la presencia de datos faltantes. Su funcionamiento no se realiza adecuadamente cuando hay estos espacios vacíos (Graham, 2009; Marwala, 2009; Kumar, 2022). El primer paso para abordar este problema consiste en identificar los motivos por los cuales hay valores ausentes, a fin de tener un mejor control del problema; esta información puede ser útil para determinar si las pérdidas comprometen de forma riesgosa al conjunto de datos. Tratar apropiadamente estas situaciones es más importante de lo que parece. Actualmente prevalece la tendencia de minimizar los efectos de eliminar o de sustituir los valores ausentes de la matriz de datos, buscando facilitar el procesamiento y

disfrutar de una base completa para ser analizada. Es sabido que la ausencia de unas cuantas observaciones, por muy pocas que sean, pueden influir más de lo que uno pensaría, provocando que los resultados de los análisis se debiliten o se sesguen (Soley-Bori, 2013). Useche y Mesa (2006) indican que las consecuencias de la “no respuesta” se traducen en resultados estadísticos poco eficientes, distorsión en las frecuencias, estimaciones sesgadas, menor tamaño de muestra y la falta de representación en las variables.

En esta encrucijada es necesario que quien investiga ponga en tela de juicio aspectos como la proporción o cantidad permisible de datos faltantes (en función del problema que se está abordando). Debe tener siempre en mente si estos se deben simplemente a la aleatoriedad, a causas de fuerza menor o a una deficiente recolección del instrumento; lo que debe conducir a tomar una decisión sobre los datos faltantes o, aún más importante, a replantear la idea de si es apropiado o no aplicar el análisis sobre los datos.

La falta de datos puede presentarse de distintas maneras. La primera de ellas es la pérdida total, que es cuando falta el registro de todas las variables de una(s) unidad(es). La segunda, la parcial, es cuando existe la ausencia en una o más variables, pero no en individuos completos. Es decir, hay algunas unidades de las que se recopilaron solamente algunas características. En los casos en los que la pérdida es parcial, como comúnmente es el caso, se puede recurrir a diversas soluciones; sobre las que se debe estar consciente de cuáles son las implicaciones de su aplicación y de las precauciones que se deben tomar.

Los recursos para abordar estos problemas fueron descritos hace ya varias décadas, en el trabajo de Rubin (1976). Desde entonces se han estudiado nuevas rutas para lograr sustituir esos datos faltantes, procurando que no se altere la naturaleza de la información. Para solucionarlo, se deberá decidir entre dos enfoques. El primero de ellos alberga técnicas cuya solución es sencilla de implementar desde un punto de vista computacional, pero que implica mayores riesgos de crear sesgos e inconsistencias en los resultados del análisis. Y que, pensándolo bien, nunca debería considerarse como la primera alternativa.

En este primer grupo de estrategias se considera eliminar por completo aquellos registros u observaciones que contengan al menos un valor ausente para alguna de las variables. Para tomar esta decisión se debe tener en mente si la pérdida de datos sería mínima o si, en caso contrario, al hacer esto se está excluyendo una gran cantidad de datos que pueden ser cruciales para la investigación. Como explican Useche y Mesa (2006), la ausencia de datos probablemente pueda ser ignorada en grandes conjuntos de datos en donde la proporción de pérdida es poca, pero esto no es conveniente cuando es el caso opuesto.

Por otra parte, se puede optar por sustituir de forma conveniente los valores ausentes de las variables a través de la deducción e intuición. Sobre todo, cuando se conoce con cierta certeza el motivo de las ausencias. Aunque existe el riesgo de llegar a una base de datos poco justificable. Preferentemente no se deben considerar estas alternativas como una primera opción, ni segunda o tercera. A menos que se tengan criterios

estrictamente sólidos y argumentos válidos que justifiquen su aplicación.

El segundo grupo de estrategias está constituido por un conjunto de técnicas que, en sintonía con el desarrollo informático, consideran en su totalidad la base de datos y no sólo de forma univariada, como lo es en el primer caso. El segundo enfoque se apropia de los casos en donde existen múltiples variables y hay valores ausentes, mediante la aplicación de métodos que imputan datos de forma conjunta.

Imputación de datos

Cuando los datos faltantes no pueden pasar desapercibidos y no es factible simplemente eliminarlos, se deben de sustituir esos espacios vacíos por otros valores. En términos generales, a esto se le conoce como imputación. Sobre este marco, se desprenden diversos procedimientos de imputación. Actualmente, el desarrollo del cómputo estadístico nos ha brindado el acceso a técnicas especializadas que han sido diseñadas con el propósito de imputar datos. A pesar de esto se siguen implementando con bastante frecuencia las técnicas de imputación más sencillas que, como ya se estableció, casi nunca son una alternativa confiable. Esto motiva a que el uso de estas técnicas especializadas se amplifique, a su vez que estas herramientas sean consideradas como recursos de primera mano cuando el usuario se encuentre en este momento de la investigación.

Imputación mediante la media o moda

Este recurso consiste en sustituir cada uno de los valores ausentes por medidas de tendencia central. La media muestral calculada a partir de los datos disponibles para el caso de variables del tipo cuantitativo, y la moda para las cualitativas. Ya es conocido que, aunque esta salida luzca razonable su uso conlleva consecuencias. Y al igual que la eliminación de casos, se ha evidenciado que su uso puede terminar agravando más el problema que solucionándolo. ¿Por qué? Pues la razón es simple: reduce la variabilidad, sobre todo cuando la cantidad de datos faltantes es grande, pues todos los datos faltantes (para una misma variable) se “rellenarían” con el mismo valor. Estas prácticas son conocidas como imputaciones simples. Desde inicios del siglo XXI, Pérez y Alonso (2002) destacaron que éstas son poco apropiadas para atacar el problema de los espacios vacíos. Estos recursos podrían ser efectivos cuando se trata de muy pocos valores ausentes.

Pero la imputación puede ir más allá de la sustitución por un valor único. Esto es llenar aquellos espacios vacíos con el valor más probable. A través de técnicas que contemplan la naturaleza multivariada del problema y que al mismo tiempo “recuperan” estos datos auxiliándose de la mayor cantidad de información. Al emplear este tipo de imputaciones no se altera de forma significativa la variabilidad de los datos, dado que a cada valor ausente le es asignado un valor en función de los principios y fundamentos matemáticos y estadísticos que rigen cada una de las técnicas.

Por lo tanto, se cuenta con un conjunto de técnicas de mayor eficiencia, respaldadas por literatura reciente (y

que por supuesto siguen surgiendo modificaciones e innovaciones en este rubro). No obstante, en la medida en que se aspira a una mayor eficiencia y calidad de los datos, se exige una mayor apropiación de la programación estadística y del manejo de software. Por su parte, estas técnicas van impregnadas de criterios y metodologías muy específicas que ayudan a quien investiga a determinar qué tan apropiado es su uso, según sea el caso. Por ejemplo, hoy en día, se sabe que la imputación múltiple es una de las técnicas más eficientes para resolver los problemas de datos faltantes, demostrando mejores resultados en las inferencias estadísticas después de ser aplicada. Cabe destacar que esta solución incorpora un factor de incertidumbre. Como bien describen Bacallao y Bacallao (2010), la imputación múltiple se encarga de completar la base de m formas diferentes, creando distintos conjuntos de datos, los cuales se someten al análisis estadístico que se había pensado inicialmente, obteniendo distintas estimaciones y errores, combinados para encontrar una solución única. Sin embargo, requiere de varios supuestos para su uso y su implementación no se extiende a los datos del tipo cualitativo. El sustento teórico de este recurso y su aplicación se detalla en otros trabajos (Dong y Joanne, 2013; Mejía-Giraldo y Restrepo-Betancur, 2019).

Así mismo, existen otros métodos estadísticos de imputación, como el algoritmo *Expectation-Maximization* (Dempster et al., 1977; Randahl, 2016), el modelo de covarianza para datos experimentales (Cochran, 1957; John y Prescott, 1974) y, más recientemente, los árboles de decisión y los bosques aleatorios (Hong y Lynn, 2020; Tang, 2017), por mencionar algunas de las más populares. Especialmente, las técnicas de minería, como

los bosques aleatorios, se destacan por dos poderosas razones: son capaces de imputar variables cualitativas y cuantitativas de forma simultánea, además de que su rendimiento y calidad no se ve afectada ante grandes conjuntos de datos; como sí puede pasar con la imputación múltiple y el algoritmo EM.

Ahora bien, ¿cómo puede quien investiga saber qué tan grave es el problema? ¿Sobre qué aspectos debe prestar atención de los datos faltantes? Se puede hacer un diagnóstico del problema, evaluando los siguientes aspectos:

1. Patrón de pérdida: ¿Las pérdidas se dan de forma aleatoria? ¿Qué combinaciones de variables están perdidas para las mismas unidades? ¿Las pérdidas en una variable se le pueden atribuir a otras variables que sí se observaron? Si tal parece que las pérdidas se dieron de manera aleatoria y estas pérdidas no se ven influenciadas por otras variables, el problema puede no ser grave.
2. Diferencia entre casos: ¿Existen diferencias entre las unidades que sí tienen datos de las que no? ¿Son estadísticamente diferente los individuos que sí contestaron de los que no? ¿Hay categorías de alguna variable que alberguen una mayor cantidad de datos faltantes? Por ejemplo, si se tiene registrado el sexo de 100 estudiantes, y también se les pregunta que opinan del feminismo (siendo este un tema recurrente) es probable que haya un sexo con mayor inclinación a no contestar; y si la encuesta trata principalmente de este tema, como en nuestro ejemplo ficticio, muy probablemente habrá una diferencia entre los casos observados y los no observados. Por otra parte, si la distribución de variables continuas es distinta entre los casos observados y los no observados; por supuesto

que la situación puede volverse más complicada al evaluar esta situación con múltiples variables. Si los datos tienen este problema de la diferencia entre casos, el problema puede ser grave y las técnicas sencillas poco pueden hacer para mejorar esta situación.

3. Tasa de perdidas: Por cada variable (columna), por cada unidad de estudio (fila) y en su totalidad ¿Qué porcentaje representan los datos faltantes en el conjunto de datos? ¿Es este porcentaje significativo? Luengo et al. (2010) proponen mirar este aspecto de la siguiente forma: 1 % de faltantes es trivial, entre 1 % y 5 % puede ser manejable, entre 5 y 15 % sin duda requiere de técnicas especializadas, y faltantes por arriba de 15 % tienen una alta probabilidad de comprometer los resultados, aun así, pueden ser tratadas. Vale la pena, resaltar qué si los datos faltantes se deben a una causa únicamente de la aleatoriedad, la gravedad de los faltantes es menor.

Normalización de los datos

La normalización de los datos es una técnica que transforma la escala de las variables de interés, buscando la posibilidad de hacer comparaciones entre distintos elementos y escalas de medición. Generalmente esto se emplea cuando se requieren calcular las probabilidades de los valores dentro de una variable y cuando se conoce de antemano que la distribución de la variable se aproxima a una normal. Como se describe en Jiawei et al. (2012), las unidades de medida pueden impactar sobre el análisis de los datos y llevar a otros resultados; para eludir apropiadamente la influencia de la diferencia en las unidades de medición, los datos deben normalizarse o estandarizarse. Esto quiere decir, que los atributos o variables pasarán a un

...en el proceso de investigación

rango más pequeño, por ejemplo: [-1, 1] o [0 a 1]. Desde el enfoque estadístico, esta transformación permite darles el mismo “peso” a las variables.

Por otra parte, el uso de transformaciones es necesaria para poder aplicar algunas de las técnicas estadísticas multivariadas más utilizadas en las investigaciones del corte social como: el análisis de componentes principales, el análisis factorial confirmatorio/exploratorio, así como en algunas de las técnicas de aprendizaje máquina y minería de datos, sobre todo en las técnicas de clasificación. Existen diversos métodos para hacer la normalización de los datos. Vale la pena aclarar que algunos de estos métodos se incorporan de forma automática dentro de las rutinas de los análisis que requieren este tipo de transformaciones.

Categorización y colapsamiento de variables

En ocasiones conviene categorizar y/o colapsar algunas variables. Se debe considerar el tipo de estudio, la estructura de los datos, los propósitos de la investigación y la selección tentativa de las técnicas para los análisis definitivos. La primera técnica consiste en transitar de una variable numérica a una nueva variable categórica construida a partir de la original. Estos valores numéricos son remplazados por intervalos (60-75, 76-90, etc.) o mediante conceptos específicos (bajo, medio, alto). Esto permite simplificar en cierta medida el conjunto de datos. Como en este ejemplo, la aplicación de este recurso resulta útil cuando una variable cuantitativa puede ser expresada en forma de intervalos mediante categorías

en concreto. Existen dos métodos, los cuáles son los más comunes para categorizar.

Categorías con el mismo ancho

En este caso el conjunto de datos se divide en k intervalos que tengan el mismo tamaño. Es decir, si se cuenta con datos que oscilan en el rango de 1 a 30, se podrían generar tres categorías con el mismo ancho (1-10, 11-20, 21-30). Para esto se utiliza la siguiente fórmula:

$$w = \frac{\max - \min}{k}$$

Categorías con la misma frecuencia

Este recurso se encarga de dividir el conjunto de datos en k grupos, de tal manera que cada uno de esos grupos o categorías tengan la misma cantidad de valores. Para escoger el número de grupos o el número de intervalos adecuado es necesario que se observe cómo están distribuidos los datos mediante un histograma; determinando la forma que mejor convendría para agrupar los datos.

La segunda técnica mencionada consiste esencialmente en reducir los niveles de una variable que ya es categórica. Colapsar variables en estadística implica hacer nuevas agrupaciones entre las respuestas de una misma variable, buscando obtener grupos con mayor representatividad. Al contar con grupos más grandes de respuestas, las pruebas estadísticas tienden a ser más confiables. Esto permite generar un discurso

...en el proceso de investigación

con resultados y hallazgos estadísticamente significativos. A continuación, se muestra un ejemplo de ello.

Tabla 5. *Colapsamiento de variables.*

Variable original	Variable colapsada
Baja California	Norte
Veracruz	Sur
Oaxaca	Sur
Jalisco	Centro
Michoacán	Centro
Chihuahua	Norte

Nota: elaboración propia.

Exploración inicial o análisis exploratorio

Resulta útil hacer una exploración de los datos antes de comenzar con los análisis definitivos. La exploración inicial, o análisis iniciales, constituye una fase en la que se busca tener el primer acercamiento a la información que hay en los datos. Este proceso básicamente se reduce a la descripción de los datos a partir de gráficos acompañados de estadísticas básicas. El análisis exploratorio generalmente no es explicado ni visualizado en un reporte final. Sin embargo, esta exploración inicial debe realizarse y es indispensable para pasar a los análisis de profundidad, ya que los resultados permiten comprender cómo están distribuidos los datos, además de cuantificar y observar su comportamiento; así mismo suman información útil para escoger la técnica de análisis definitivo más adecuada.

Es aquí donde se caracterizan esencialmente dos aspectos básicos de las variables: tendencia central y dispersión. Para ambas partes, con apoyo del software

estadístico, calcular estas estadísticas descriptivas se ha convertido en una tarea que requiere de poco esfuerzo. La media, la mediana o la moda son las estadísticas más utilizadas para describir al individuo/elemento “típico”; la varianza, la desviación estándar y los cuantiles ayudan a cuantificar qué tanto varían los datos alrededor de este dato “típico”. La exploración de las variables categóricas se hace mirando directamente sus frecuencias o porcentajes en acompañamiento de un gráfico que las ilustre. Posterior a ello se recomienda usar elementos gráficos que permitan visualizar esta información. Esta exploración previa permite identificar los valores faltantes, posibles incoherencias, inconsistencias o posibles errores en la base de datos. Ahora el software estadístico consigue que esta etapa se lleve a cabo fácilmente al realizar despliegues gráficos y calcular las estadísticas descriptivas de forma automática, con unos cuantos comandos o *clics*. A continuación, se explica cómo estas visualizaciones y medidas aportan información útil para que quien investiga tome decisiones respecto al procesamiento definitivo de sus datos.

Tendencia central de los datos

Aunque la media, la mediana y la moda nos aporten ideas sobre el valor central de la distribución de los datos, no son intercambiables entre sí. La media (o valor promedio) es la medida de tendencia central más utilizada para describir a las variables numéricas. Sin embargo, no siempre se considera la más apropiada. Como se sabe, la media, es sensible a los valores atípicos o datos extremos; incluso aunque sean pocos pueden influirla de forma significativa. En dichos casos

conviene examinar con más detalle cuáles son estos datos atípicos y si su naturaleza se debe a un error de medición o si realmente constituye un valor real. En el segundo caso es factible utilizar otra medida de tendencia central como lo es la mediana. La mediana, a diferencia de la media, no sufre de este problema y su interpretación es igualmente intuitiva. La mediana representa el punto que separa al 50 % más alto, del 50 % más bajo. Finalmente, la moda se utiliza principalmente para los datos del tipo cualitativo o para variables discretas (número de hijos, años de experiencia laboral, etc.), aunque su uso también se puede extender a las numéricas. Este valor es definido simplemente como el valor más común. Las medidas de tendencia central pueden ser obtenidas de manera fácil con los paquetes estadísticos. Pero, además del cálculo de las medidas de resumen, es necesario visualizar la forma en la que se distribuyen los datos. Esto permite obtener un panorama con mayor claridad acerca de los datos en conjunto, así como un criterio para escoger la medida que se usará para describir a la variable. Por tal motivo se recomienda primero visualizar la distribución de los datos y posteriormente se entenderá mejor su resumen a través de las estadísticas descriptivas.

Dispersión y variabilidad de los datos

Como se mencionó anteriormente, para cuantificar qué tan dispersos se encuentran los datos se utilizan los valores de la varianza, la desviación estándar y los cuantiles. Sin embargo, estas medidas sólo pueden ser empleadas para variables cuantitativas. Por ejemplo, cuando el valor de la desviación estándar es bajo,

significa que nuestros datos se encuentran muy cercanos a los valores centrales de la media o la mediana; de lo contrario indicaría que los datos se distribuyen de manera más dispersa dentro de un rango de valores de mayor amplitud.

Por otro lado, los cuartiles aportan información de una manera distinta. Estos pueden ser calculados a través de cualquier software estadístico, o mediante los gráficos de cajas y alambres. Generalmente se designan como Q1, Q2 y Q3. El primero de ellos es el primer cuartil e indica el punto bajo el cual se encuentra el 25 % de los datos. Es decir, si se tienen los datos de la edad de 150 trabajadores y el Q1 tiene un valor de 23 años, esto indicaría que el 25 % de los trabajadores tienen 23 años o menos. El Q2, es el equivalente a la mediana. Y finalmente, si el Q3 tiene un valor de 67 años, indicaría que el 25 % de los trabajadores tiene más de 67 años o que el 75 % de ellos tiene 67 años o menos. Su interpretación es sencilla, pero sumamente útil. Aportan a quién investiga información relevante acerca de cómo se distribuyen sus variables, respecto a su distribución en una partición del rango en cuatro partes; también se utilizan para evaluar la simetría de la distribución de los datos en el rango de valores. Como se dijo, los gráficos de cajas y alambres muestran gráficamente este aspecto, sumamente importante en el análisis de los datos.

Primera visualización de los datos

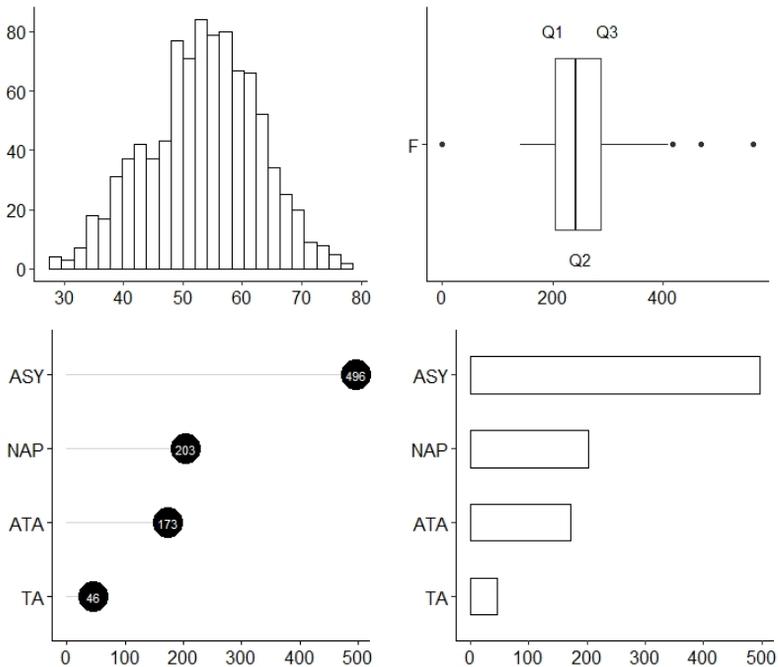
Para ello existen una serie de gráficos que permiten visualizar los datos en conjunto, tanto de forma univariada, bivariada como multivariada. Este quehacer

comienza ilustrando cada variable de forma individual. Para ello existen diversos gráficos, desde los más elementales para visualizar la información variable a variable. Estos son los histogramas, los gráficos de cajas y alambres, los gráficos de barras o columnas y los gráficos de puntos.

Cabe agregar que el uso de gráficos de pastel o de dona ya no es recomendable. Los gráficos deben mantener una esencia simple, tratando que la información que se quiere presentar sea clara y directa. Adicionalmente, Good y Hardin (2012) sugieren evitar los gráficos con un diseño 3D; indicando que agregar más dimensiones de forma innecesaria obstruye la comprensión del gráfico, además de que no aporta ningún tipo de información. Como mencionan Kenett y Thyregod (2006), los gráficos deben mantenerse simples evitando sobrecargarlos de elementos que no aporten información. El abanico de opciones que ofrecen los programas estadísticos para la elaboración de gráficas puede tentar a quién investiga a elaborar gráficos con detalles que entorpezcan su interpretación y distraigan la atención de lo que realmente se trata de comunicar. Los gráficos exploratorios deben rescatar los patrones de tendencia y variabilidad de las variables, en conjunto con la forma de la distribución.

Cuando se quieren presentar gráficamente las variables cuantitativas habitualmente se utilizan los histogramas y los gráficos de cajas y alambres. Permiten ver de forma completa cómo se distribuyen los datos. Para las variables del tipo cualitativo se utilizan gráficos de barras o de puntos. Se pueden utilizar tanto sus frecuencias absolutas como los porcentajes de cada categoría o valor.

Figura 5. Exploración inicial.



Nota: elaboración propia.

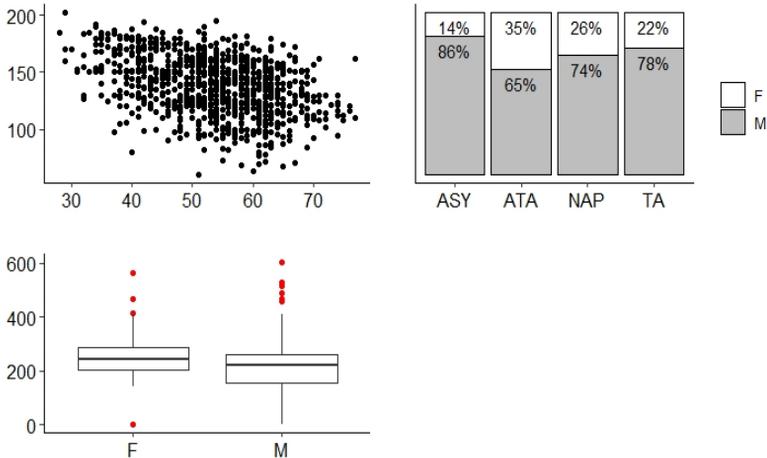
Como se mencionó anteriormente, el gráfico de cajas y alambres ilustra los valores de los cuartiles en la distribución, así como los valores máximos y mínimos, considerando el comportamiento típico del colectivo. Además, este gráfico permite identificar con facilidad la presencia de datos atípicos. Por otra parte, para las variables cualitativas se extiende el uso de los gráficos de barras o de puntos; con ellos se observa cómo están distribuidos los niveles o categorías de una variable categórica. Es relevante mencionar que en ocasiones conviene ordenar estas barras, ya que permite visualizar

con más precisión cuáles son las categorías más frecuentes, principalmente cuando el número de categorías es alto. Sin embargo, en el caso de las variables ordinales, se recomienda dejar estas barras en el orden que le corresponde de acuerdo a sus niveles. Por ejemplo, si se hizo una pregunta relacionada con la satisfacción al cliente con respuestas en una escala ordinal (Muy satisfecho, Satisfecho, Medianamente satisfecho, Poco satisfecho y Nada satisfecho), no sería adecuado cambiar su orden en función de las frecuencias en el gráfico; es decir, las categorías conservarían su lugar. A través de la exploración inicial se termina por conocer la estructura de todas las variables, tanto las de medición como las de agrupación.

Para dar otro paso en la estrategia de la exploración, se elaboran gráficos de forma bivariada. Generalmente es de mayor interés visualizar cómo se comportan las variables de medición con respecto a las variables de agrupación y cómo varían de acuerdo con sus niveles o categorías. Por ejemplo, si como variables de agrupación se tiene el sexo, y como de medición el tipo de enfermedad o la presión sanguínea, sería importante observar cómo varía el comportamiento de éstas en función del sexo. Al hacer la comparación entre dos variables se puede llegar a diferentes escenarios: dos variables cuantitativas, dos cualitativas o una cuantitativa y una cualitativa. Para el primero de los casos, lo más convencional es utilizar gráficos de dispersión. Por otro lado, cuando se tienen dos variables cualitativas, resulta práctico visualizarlas mediante los gráficos de barras apiladas o tablas de contingencia. Finalmente, en el último de los casos, se pueden utilizar los gráficos de exploración inicial pero ahora agregando la variable que

distingue a los grupos o niveles, es decir la variable categórica.

Figura 6. Exploración bivalente.



Nota: Elaboración propia.

El primer gráfico refleja el comportamiento de dos variables cuantitativas. En este ejemplo se tomaron las edades y ritmo cardíaco de 918 pacientes. Se observa que a medida que el paciente es mayor, su ritmo cardíaco tiende a ser menor; en cambio pacientes más jóvenes tienen ritmos cardíacos más altos. El gráfico de barras apiladas permite observar cómo varían las categorías entre dos variables cualitativas. En este caso, se observa cómo se distribuyen 4 tipos de dolores en función del sexo del paciente. Las tablas de contingencia son su equivalente; sin embargo, el gráfico resulta ser más ilustrativo y eficaz. Finalmente, el gráfico de cajas y alambres representa las distribuciones del nivel de colesterol, haciendo la distinción por el sexo del paciente.

...en el proceso de investigación

Preparar correctamente los datos implica corregir sus inconsistencias, posibles errores de medición y realizar cambios en las escalas o en la naturaleza de las variables. El propósito es comprender cómo están estructurados nuestros datos, así como manejarlos de manera eficiente para su exploración inicial y los análisis avanzados. Evidentemente la implementación de estos recursos y herramientas exige de la apropiación en el uso de un software especializado, así como de una mayor compenetración con los fundamentos estadísticos. Esta preparación de los datos concluye con la exploración inicial. Las estadísticas descriptivas y los elementos gráficos comienzan a vislumbrar información relevante acerca de nuestro conjunto de datos.

Conclusiones

El análisis de las evidencias, de los datos, constituye uno de los momentos más interesantes y llamativos del proceso investigativo. No obstante, llegar a ese punto requiere de un tratamiento y un pre-procesamiento de los datos que se ocuparán en dicho análisis. Lo común es que los datos recolectados requieran de un refinamiento y de un tratamiento en cada una de sus variables, antes de que sean analizadas. Los datos deben ser de la más alta calidad posible, si se quieren obtener resultados en esta misma medida.

A lo largo del capítulo se mencionaron algunas de las consideraciones que debe tener quien investiga al momento de llevar a cabo esta tarea. Vale la pena destacar que el pre-procesamiento de los datos conlleva más tiempo de lo que usualmente se cree. Para ello, quien investiga debe estar en sintonía con el manejo de

herramientas básicas de la estadística, que le permitan hacer estos ajustes sobre su base de datos. Siempre se debe tener en mente que una gran cantidad de estas decisiones están alineadas con los análisis que se tienen pensados. Por lo que es importante tener claridad de cómo están estructurados los datos y cómo se analizarán en función de sus cualidades.

En el ámbito de la investigación educativa, por lo general se recolecta y se trabaja con variables categóricas o cualitativas en combinación con variables cuantitativas. Por ello, se recomienda poner especial atención en cómo depurarlas, y cómo hacer visualizaciones más apropiadas de estas variables. Sin olvidar que los datos faltantes no deben ser tratados a la ligera y que, efectivamente, requieren un tratamiento estadístico especial. Las transformaciones que se hagan en los datos impactarán en los resultados que se obtengan; a veces en mayor o menor medida, pero lo harán.

De la misma manera, existen ciertas recomendaciones y reglas que generalmente se pasan por alto acerca de la visualización básica de la información. Como se sabe, las formas en las que se visualizan los datos han ido evolucionando. Las presentadas en este capítulo constituyen las formas básicas de hacerlo. Sin embargo, quien investiga puede compenetrarse con otras alternativas para visualizar mejor sus datos; o bien contratar a un estadístico para que le apoye en esta tarea, lo cual no lo eximirá de estar al tanto del proceso porque se requerirá de la toma de decisiones, la cual se debe hacer en sintonía con los principios estadísticos, pero sin perder de vista los objetivos y las preguntas de investigación.

Referencias

- Jorge Bacallao Guerra y Jorge Bacallao Gallestey, “Imputación múltiple en variables categóricas usando data augmentation y árboles de clasificación”, *Revista de investigación Operacional*, núm. 2, 2010, pp. 133-139.
- William Cochran, “Analysis of covariance: Its nature and uses”, *Biometrics*, núm. 3, septiembre de 1957, pp. 261-281. <https://doi.org/10.2307/2527916>
- Arthur Dempster, Nan Laird & Donald Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society*, núm. 1, 1977, pp. 1-38.
- Yiran Dong, & Chao-Ying Joanne Peng, “Principled missing data methods for researchers”, *Springer Open Journal*, núm 222, mayo 2013, pp. 1-17. <https://doi.org/10.1186/2193-1801-2-222>
- Philip Good & James Hardin, *Common errors in statistics (And how to avoid them)*, Estados Unidos, John Wiley & Sons, Inc, 2012.
- John Graham, “Missing data analysis: Making it work in the real World”, *Annual Review of Psychology*, núm. 60, enero de 2009, pp. 549-576.
- Shangzhi Hong & Henry Lynn, “Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction”, *BMC Medical Research Methodology*, núm. 199, julio de 2020, pp. 1-12. <https://doi.org/10.1186/s12874-020-01080-1>

- John & Phillip Prescott, "Estimating missing values in experiments", *Journal of the Royal Statistical Society*, núm. 2, octubre de 1974, pp. 190-192.
- Han Jiawei, Micheline Kamber & Jian Pei, *Data mining: Concepts and techniques*, Estados Unidos, Morgan Kaufmann, 2012.
- Ron Kenett & Poul Thyregod, "Aspects of statistical consulting not taught by academia", *Statistica Neerlandica*, núm. 3, 2006, pp. 396-411.
- Handling missing values with random forest. (Suni Kumar, 2022). Consultado en:
<https://www.analyticsvidhya.com/blog/2022/05/handling-missing-values-with-random-forest/>
- Susana Lloret-Segura, Adoración Ferreres-Traver, Ana Hernández-Baeza y Inés Tomás-Marco, "El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada", *Anales de psicología*, núm. 3, agosto de 2014, pp. 1151-1169.
<http://dx.doi.org/10.6018/analesps.30.3.199361>
- Julian Luengo, Salvador García & Francisco Herrera, "A study on the use of imputation methods for experimentation with Radial Basis Function Network classifiers handling missing attribute values: The good synergy between RBFN's and EventCovering method", núm. 23, abril de 2010, *Neural Networks*, pp. 406-418.
<https://doi.org/10.1016/j.neunet.2009.11.014>
- Tshilidzi Marwala, *Computational intelligence for missing data imputation, estimation and management: Knowledge optimization techniques*, Estados Unidos, IGI Global, 2009.

- Luis Miguel Mejía-Giraldo, y Luis Fernando Restrepo-Betancur, “Imputaciones múltiples, herramienta para la estimación de datos faltantes en la modelación de regresión”, *Temas agrarios*, núm. 1, julio de 2019, pp. 66-73.
<https://doi.org/10.21897/rta.v24i1.1780>
- Mario Miguel Ojeda Ramírez, “Análisis de datos”, *La Ciencia y el Hombre*, núm. 1, 1988, pp. 121-133.
- Adriana Pérez, y Martín Alonso, “Imputación múltiple, un novedoso método para el tratamiento de datos faltantes”, *Acta Médica Colombiana*, núm. 4, julio-agosto de 2002, pp. 204-208.
- David Randahl, *Raoul: An R-package for handling missing data*, [Tesis de máster, Uppsala University], 2016.
- Donald Rubin, “Inference and missing data”, *Biometrika*, núm. 3, diciembre de 1976, pp. 581-592.
- Dealing with missing data: Key assumptions and methods for applied analysis. (Marina Soley-Bori, 2013). Boston University School of Public Health. Consultado en:
https://www.researchgate.net/publication/323266125_Dealing_with_missing_data_key_assumptions_and_methods_for_applied_analysis
- Fei Tang, *Random Forest missing data approaches*, Estados Unidos, University of Miami, 2017.
- Lelly Useche y Dulce Mesa, “Una introducción a la imputación de valores perdidos”, *Terra*, núm. 31, 2006, pp. 127-152.
- Shichao Zhang, Chengqi Zhang, & Qiang Yang, “Data preparation for data mining”, *Applied Artificial Intelligence*, núm. 17, noviembre de 2010, pp. 375-381. <https://doi.org/10.1080/713827180>

Análisis definitivos: análisis multivariantes y modelación estadística

GERARDO YORHENDI CEBALLOS MARÍN
ARIEL ANTONIO LÓPEZ SALAS

Introducción

En los últimos años, el avance informático y los desarrollos computacionales han impulsado el uso de programas para aplicar los métodos estadísticos. El software estadístico, ahora, es capaz de procesar volúmenes de datos más grandes en menor tiempo. Llegando así, al punto en el que cualquier problema que involucre datos, aun siendo complejo, pueda extraerse conocimiento en un tiempo corto. Los métodos estadísticos tradicionales siguen siendo ampliamente utilizados; comprender sus principios y supuestos es clave para poder aplicarlos apropiadamente en el marco de la investigación.

En primera instancia habría que decir que la mayoría de los problemas reales son de naturaleza multivariante; es decir, en los fenómenos de estudio concurren muchos factores, los cuales deben ser considerados conjuntamente para dotar de realismo a las investigaciones. Desde la década de los sesenta del siglo pasado, ya se contaba con paquetes estadísticos para realizar análisis multivariantes, cosa que ha permitido una amplia diseminación de estas técnicas.

Las técnicas estadísticas multivariantes muchas veces se utilizan como procedimientos exploratorios, para estudiar las asociaciones entre las variables y así mismo para explorar la agrupación de los individuos. En este sentido, la creación de indicadores por medio de la reducción de dimensionalidad es una de las técnicas más socorridas dentro del análisis multivariante. Los análisis definitivos pueden someterse únicamente al uso de técnicas exploratorias multivariantes, pero con bastante frecuencia se tiene la necesidad de ajustar modelos estadísticos.

Por otro lado, es justamente por los grandes avances del cómputo estadístico que la modelación puede considerarse ya como un área de estudio y especialización. Unifica aspectos teóricos, metodológicos y computacionales. También postula, ajusta y evalúa la capacidad y la sensibilidad de un modelo para describir una relación causa-efecto alrededor de un colectivo de unidades de estudio. Cabe resaltar las palabras de Ojeda (2004), quien menciona que los modelos estadísticos han sido ampliamente utilizados para resolver problemas concretos de la ingeniería y de las diferentes áreas científicas. Esto ha permitido que se constituyan en la base de la formulación teórica de la inferencia y de muchos otros métodos. El mismo autor refiere que la aparición de la revista *Statistical Modelling: An International Journal* en el 2001, marcó un parteaguas en el desarrollo de los modelos para la resolución de diversos problemas de investigación.

En este sentido Castro y Lizasoain (2012) han referido que uno de los campos del conocimiento donde las herramientas estadísticas han servido en gran medida, ha sido el educativo. Han permitido a quien investiga en

esta área plantear modelos integrados para estudiar relaciones entre constructos no directamente observables dentro de los complejos contextos escolares. Para ello hacen uso de árboles estadísticos de decisión, modelos de ecuaciones estructurales y modelos multinivel, por mencionar algunos. Barreto-Villanueva (2012) describe que la metodología estadística ha auxiliado a la investigación pedagógica empírica, contribuyendo significativamente a “la comprensión de los métodos de investigación en educación, el diseño de programas, los problemas de medición y evaluación, el diagnóstico y hasta su orientación” (p. 6).

En cuanto a las investigaciones en ciencias sociales (CS), los modelos estadísticos han servido para aproximarse y estimar el desarrollo de diferentes comportamientos colectivos, la descripción de procesos institucionales y sociales, la evaluación de organizaciones y su interrelación con diversos factores del entorno. De igual forma para el análisis y la comparación de las estructuras sociales subyacentes, así como sus repercusiones en diferentes contextos.

Desde que las CS se comenzaron a estudiar ha existido una pugna entre dos vertientes metodológicas: la cualitativa, que prioriza la recolección de valores objetivos medibles mediante técnicas como la observación, la entrevista, la participación grupal y los grupos focales, por ejemplo; y la cuantitativa, que se centraliza en la recopilación y análisis de datos numéricos utilizando exámenes, escalas de valoración y encuestas. Hoy en día, gracias a los métodos mixtos, pueden combinarse ambas metodologías empleando preceptos estadísticos en los instrumentos que

tradicionalmente han sido de investigación cualitativa, como la entrevista y los grupos focales (Barreto-Villanueva, 2012).

Debe resaltarse que, si bien las técnicas cuantitativas siguen siendo ampliamente usadas, en las últimas décadas ha habido una tendencia hacia el incremento en el uso de métodos mixtos. Un ejemplo la iniciativa *Q-Squared*. Esta propuesta engloba a una serie de iniciativas de investigación que han utilizado el método mixto o Q-cuadrado (Q2), con enfoques cualitativos y cuantitativos combinados para llevar a cabo el análisis de la pobreza. En ellas se destaca la amplia gama de herramientas analíticas dentro de las CS que pueden usarse para comprender y explicar los fenómenos sociales. Sobre todo, al momento de investigar cuestiones fundamentales en torno a los debates sobre la pobreza con respecto a la epistemología y la causalidad (Schaffer, 2013). Sin embargo, a pesar de esta creciente popularidad metodológica, los métodos mixtos todavía representan tan sólo una parte diminuta del trabajo publicado en muchas naciones en desarrollo (Harris, 2021).

Retomando este marco, es importante que los procedimientos estadísticos multivariados y los modelos estadísticos sean reconocidos como herramientas fundamentales en el desarrollo profesional de quienes son especialistas en educación y CS. Suelen llevar asignaturas y temas selectos de estadística, pero muchas veces no llegan a comprender ni poder aplicar correctamente los conceptos y procedimientos subyacentes. Por tal motivo, en las siguientes páginas se presentan algunas técnicas multivariantes y de

modelación estadística. Se destacan las características y se mencionan las potenciales aplicaciones.

Primeramente, debe establecerse una definición básica sobre lo que es un modelo matemático y para qué sirve en los distintos ámbitos científicos. Este es un constructo con el que se puede estudiar y comprender de mejor forma un fenómeno del que subyace una relación causa-efecto, del tipo $X \rightarrow Y$. Por lo tanto, este modelo expresa un postulado acerca de la naturaleza de esta relación, el cual se denota a través de una formulación matemática.

Derivado de él surge el modelo estadístico. Este considera la medición sobre una variable explicativa, una unidad de estudio u_i , la cual produce una observación $y_i = Y(u_i)$. Dicha observación se concibe en dos componentes, sean “señal o ruido” o un patrón o simple dispersión. Ambos se denominan genéricamente la parte sistemática —la función $f(x)$ que implica una relación entre variables— y la parte aleatoria —el índice e , que señala el margen de error debido a variables no observables— (Ojeda, 2004).

Bajo estos lineamientos conviene aclarar que los datos (y más los obtenidos en estudios pedagógico-sociales) no suelen ajustarse perfectamente a un modelo teórico. Por ello es trascendental contar con una bondad de ajuste o un estudio cuidadoso de los residuos de este modelo. A su vez, debe existir un estudio de la variabilidad de los datos, tanto dentro del modelo como en los residuos. De esta forma se facilita la búsqueda de explicaciones causales de las desviaciones y varianzas entre grupos. Así, es posible realizar predicciones,

controlar procesos y determinar factores de error (Batanero et al., 2013).

Efectuando una revisión histórica abreviada, cabe advertir que en la metodología para la elaboración de modelos estadísticos se ha recapitulado que empezó con la estimación de mínimos cuadrados de Gauss en 1809. No obstante, fue el descubrimiento de la Ley de regresión filial, por parte de Galton alrededor de 1890, las aportaciones de Markov en 1900, y la subsecuente teorización matemática elaborada por Karl Pearson en el periodo 1900-1910, los que dieron origen a los modelos de regresión lineal (MRL) (Ojeda, 2004).

Otra aportación relevante fue la de Fisher a principios del siglo XX. Desarrolló de forma integral una familia de modelos para resolver un tipo genérico de problemas, dando a luz el análisis de la varianza y los correspondientes modelos de análisis de varianza (ANOVA). A su vez, Bartlett en 1935 publicó un trabajo para modelar tablas de contingencia con base en datos discretos; no obstante, fue hasta los años cincuenta cuando Lancaster, Roy y Kastenbaun diseñaron los modelos log-lineales, así como Bhapkar, Koch, Grizzle y Starmer generaron modelos lineales generales para datos en tablas de contingencia. Sumado a ello, Birch en 1963 expresó el modelo log-lineal en la forma actual, equivalente a los modelos ANOVA (Ojeda-Ramírez y Gutiérrez-Reyes, 2005).

El detonante de la modelación estadística en datos discretos lo constituyó el trabajo de Nelder y Wedderburn (1972). Tomando como base los modelos lineales generalizados, un marco teórico general para el estudio de los modelos estadísticos. En él incluyeron aquellos basados en la regresión lineal para respuestas

continuas, dicotómicas (logística), de conteos (Poisson) y los modelos ANOVA. En esta misma época se comenzó a desarrollar el análisis multivariante, el cual ha contribuido para el desarrollo de estudios tanto de dependencia como de interdependencia entre variables. Su uso se extendió y creció vertiginosamente a partir de los años 70; esto se debió a cuatro factores (Closas et al., 2013):

- En muchas investigaciones científicas se requiere analizar relaciones simultáneas entre tres o más variables.
- Estos análisis ayudan a la modificación de las hipótesis, proceso que permite que se añadan y eliminen continuamente variables.
- Entre mayor sea la complejidad de los fenómenos estimados, las variables implicadas serán más numerosas, lo que vuelve a los análisis necesariamente multivariantes.
- El desarrollo de TIC con mayor capacidad de almacenamiento, una potencia de procesamiento suficiente y programas estadísticos accesibles y con mayores herramientas para la comparación de datos (p. 66).

Dentro del análisis estadístico se pueden distinguir dos grandes tipos de técnicas. Las primeras son las explicativas o de dependencia. Están centradas en la existencia o ausencia de relaciones entre dos grupos de variables dependientes e independientes, donde las primeras son influenciadas por las segundas de manera conjunta o individualmente. Ejemplos de ellas son la regresión lineal múltiple (RLM), el análisis discriminante y el análisis multivariante de la varianza (MANOVA).

...en el proceso de investigación

Las segundas técnicas son descriptivas o de interdependencia. Se emplean en caso de que en un estudio sea imposible distinguir conceptualmente las variables dependientes e independientes. Consecuentemente se vuelve imperativo determinar cómo y por qué las variables están correlacionadas entre sí, su interdependencia (Closas et al., 2013).

Modelación estadística

Dentro de la modelación estadística se agrupan esencialmente las técnicas que asumen dependencia o la explicación de una variable a partir de otras. Es decir, por lo general existe conceptualmente una variable dependiente o respuesta, y una o más variables independientes que explican el comportamiento de la dependiente. Al abordar esta multiplicidad de esquemas han surgido técnicas de regresión en función del tipo de variables y de la forma funcional calculada entre ellas. Las más elementales (aunque las más potentes e informativas) son las lineales, pues permiten comprobar la relación entre dos variables que pueden ser transformadas de forma lineal (Montero, 2016).

A pesar de que los modelos lineales son ideales para verificar interacciones entre variables, no se suelen ajustar a las condiciones diversas y multifactoriales de los estudios sociales y educativos, entre muchos otros. Es por ello que también ha surgido la modelación para muestras complejas, cuyo origen proviene de los modelos de efectos mixtos aplicados en genética animal (propuestos por Henderson a partir de 1975). Hay que tomar en cuenta que previamente se había constituido su formulación teórica en el marco de la denominada

inferencia bayesiana. Tal inferencia se basa en el teorema de Bayes. Indica que, una vez observados los datos, toda la información sobre una variable está contenida en su distribución *a posteriori*. Dicho de otra manera, la probabilidad de hallar el evento o la característica deseada depende de un hecho que ya es conocido, el cual condiciona su observación. Sin embargo, fue hasta los años 80 del siglo XX cuando se constituyeron aplicaciones específicas y estrategias de modelación con base en la complejidad de la estructura muestral de investigaciones en entornos académicos y sociales.

Complementando lo anterior, en la década de 1990 con la expansión del Internet, el análisis estadístico se amplió con técnicas como la lógica difusa, el razonamiento heurístico y las redes neuronales. Partiendo de ellas se logró un mayor poder y una mayor eficiencia en la exploración e identificación de datos útiles, y en las formas de obtención de los mismos. Todas esas técnicas se han aprovechado para generar conocimiento mediante la minería de datos. Su enfoque está dirigido hacia métodos y algoritmos que optimizan la extracción automática de información sintetizada para caracterizar las relaciones escondidas en grandes cantidades de datos. La intención es tener capacidad predictiva, facilitando de este modo el análisis de fenómenos complejos, extensos y recurrentes de forma cada vez más eficiente (Beltrán Martínez, 2015). En este sentido, para propósitos prácticos, se distingue entre modelos paramétricos y los modelos no paramétricos.

Modelos paramétricos (MP)

Son los más conocidos y empleados, teniendo la forma general $Y_i = f(x_i; \theta) + e_i$, donde θ , es el parámetro de interés y determina el miembro particular de una familia de distribuciones indizada en un espacio paramétrico. El más importante es el Modelo Lineal General (MLG), que tiene la forma genérica: $Y_i = \alpha + \sum_{j=1}^p X_{ij} \beta_j + e_i$. Un ejemplo de él es el modelo de Regresión Lineal Simple (RLS), con fórmula $Y_i = \alpha + \beta x_i + e_i$, donde hay variables explicativas (x_1, x_2, x_p) y los β_j son los coeficientes asociados, que junto con α , son los parámetros de interés. Pues, son los que permiten comprender numéricamente como se relacionan las variables independientes con la variable respuesta. Los MP son ampliamente utilizados por diversas razones: (a) su multifuncionalidad —descripción, estimación, predicción, pronóstico y calibración—; (b) su interpretación es sencilla pero efectiva; (c) porque aproximan bien relaciones funcionales complejas; y (d) porque existe una vasta teoría y facilidades computacionales para hacer modelación estadística.

Su mayor ventaja es la posibilidad de efectuar una correlación, medida de la similitud de la variabilidad de dos magnitudes deterministas (conocidas como variables estocásticas). Aunque el cálculo de sus estadísticas es computacionalmente sencillo, tiene el inconveniente de que, en ocasiones, puede no ser suficiente para comprender la naturaleza de la relación entre variables. Quizá los modelos más conocidos en este grupo son los de la RLS y RLM, los cuales se emplean cuando se genera un conjunto de datos cuantitativos a partir de la

sustitución de valores de una variable explicativa (Ojeda, 2004). La RLS otorga información de la magnitud de la correlación, el incremento marginal entre variables y si dicha relación es considerada significativa o no en términos de linealidad.

En el caso de la RLM, lo que se trata es de ajustar modelos linealizables entre una variable dependiente y un cierto número de variables independientes. En este tipo de modelos es importante, se tiene que probar la heterocedasticidad, la multicolinealidad y la especificación, de acuerdo con Montero (2016). No obstante, las exigencias para la aplicación de este modelo son mayores cuando se trata de Modelos de Ecuaciones Estructurales (MEE). El tamaño muestral que habitualmente se requiere debe ser superior a los 100 objetos, entidades o individuos, o al menos cinco veces más observaciones que el número de variables que van a ser analizadas (aunque la tasa óptima entre observaciones y variables para seleccionar la muestra, es de diez a uno) (Closas et al., 2013).

Es interesante la acotación que realizan Closas et al. (2013) sobre las aportaciones de la RLS y la RLM. Señalan que hay evidencias contundentes, tanto matemáticas como empíricas, de que “las pruebas con una sola variable dependiente (análisis de la varianza y covarianza, de regresión, etc.) son altamente robustas bajo la violación de los principios de normalidad y homocedasticidad, excepto cuando las muestras son muy pequeñas y desiguales”. En contraste, en el análisis multivariante, “la tendencia está en considerar que en muestras grandes ($n > 30$) sus métodos son lo suficientemente robustos como para ser insensibles a

ligeras desviaciones de los supuestos estadísticos” (p. 68).

Dentro de este campo amplio de MLP, los modelos ANOVA se han empleado comúnmente cuando los datos derivados de la sustitución de valores de la variable explicativa son cualitativos, pudiendo ser aplicados en modelos de efectos fijos, aleatorios o mixtos. Sus formas utilizan las variables indicadoras para representar las categorías de las variables explicativas, ambas cualitativas.

Asimismo, los modelos de análisis de covarianza (ANCOVA) combinan el modelo de regresión lineal con el ANOVA. En ellos se asume el supuesto de que las rectas de regresión son paralelas. Cuando este supuesto no es razonable, se puede recurrir a los modelos de regresión con variables indicadoras (Ojeda, 2004).

A pesar de todas estas ventajas, uno de los problemas primordiales con los que se han enfrentado los modelos lineales es el referente a la multicolinealidad entre las variables independientes en la RLM. Por consecuencia se han aplicado métodos como la regresión Ridge, que consiste en encontrar mejores estimadores de los parámetros del modelo; otro ha sido la regresión sobre Componentes principales, técnica multivariante que sirve para corregir los estimadores de variables interrelacionadas (Del Valle Moreno y Guerra Bustillo, 2012).

Con respecto a las innovaciones, Chiok (2014) ha destacado que los modelos de redes neuronales han emergido como los nuevos paradigmas del análisis estadístico al funcionar de igual manera como algoritmos regresores. Además, estos modelos no están condicionados al cumplimiento de supuestos teóricos,

como lo es en los modelos estadísticos clásicos. De hecho, este autor refiere que el *modelo del perceptron multicapa* es equivalente a un MRL, debido a la similitud de la variable de salida (respuesta) que se relaciona aplicando la función de activación (identidad) sobre una combinación lineal de pesos (coeficientes) con las variables de entrada (predictoras). Aunque su uso se ha extendido a diferentes áreas de investigación, su uso se limita por la dificultad para elaborar una interpretación definitiva de los parámetros y de su efecto sobre la variable respuesta.

Modelos lineales generalizados (MLG)

Para el análisis de datos categóricos, los modelos log-lineales fueron clave para modelar las relaciones en tablas de contingencia. Sin embargo, la necesidad de analizar datos más complejos para las CS y económicas, llevó al surgimiento de otros modelos. Los cuáles pudiesen manejar variables dependientes no continuas. De esta forma se impulsaron modelos como los de regresión logística, regresión Poisson, modelos *probit* y *quantit*, así como los modelos lineales generales (Nelder y Wedderburn, 1972). Estos tienen la capacidad de contemplar cuando a un modelo de efectos fijos se le puede atribuir una distribución de probabilidad que es miembro de la familia exponencial (Binomial, Poisson, Gamma, Log-normal, etc.). Es decir, los MLG se caracterizan por su flexibilidad para ajustar una relación entre variables, en donde la variable respuesta o dependiente no es normal. Es decir, los MLG permiten modelar problemas en donde la respuesta de interés es

dicotómica, ordinal, continua que no siga una distribución normal o datos de conteos.

Conviene añadir que la inclusión de un efecto aleatorio en el componente sistemático es una analogía de los modelos lineales de efectos mixtos, lo que los vuelve generalizados al considerar al efecto aleatorio en la parte sistemática, suponiendo una distribución normal. Por su parte, Lee y Nelder (1996) agregaron los MLG jerárquicos y una teoría unificada de estimación y prueba de hipótesis basada en la h-verosimilitud.

Al respecto, Batanero et al. (2013) han indicado que esta clase de modelos son útiles para describir y predecir propiedades de los datos en conjunto, y no sólo de forma aislada. De tal modo puede realizarse un razonamiento distribucional que sirve para conectar los datos (mediante una distribución específica de ellos); la población de donde se tomaron (distribución de probabilidad); y las posibles muestras de la misma (distribución muestral).

A propósito de lo anterior, se ha reportado que el objetivo de las CS en cuanto a estimar, explicar, predecir y controlar niveles de variables (dicotómicas, continuas, discretas, tipo Likert, etc.) de la realidad social y sus interrelaciones, está influenciado en gran medida en el tamaño de la muestra seleccionada o en la cantidad de observaciones directas recolectadas y analizadas. Se establecen los intervalos de confianza, las aperturas robustas y la medición de las magnitudes de los efectos. En consecuencia, el número de sujetos de estudio es fundamental tanto para asegurar la pertinencia del modelo de inferencia estadístico a utilizar como para conseguir un impacto significativo en sus resultados a partir del tipo de análisis preferido (factorial, alfa de

Cronbach, regresiones, ecuaciones estructurales, etc.). Con base en ello, pueden describirse escenarios próximos de la realidad social y facilitar la toma de decisiones claves (Rositas, 2014).

Modelos no lineales (MNL) y Modelos no paramétricos (MNP)

Se requieren cuando la forma funcional $f(x_i; \theta)$ no se expresa como una combinación lineal en los parámetros. Algunos de estos son linealizables aplicando transformaciones, pero otros deben ajustarse en su forma funcional sin transformar. Cabe destacar que hay modelos paramétricos no lineales de uso amplio en ecología, ciencias biológicas y en economía, como los modelos de Gompertz, los de Weibull y los de Michaelis-Menten, entre otros. En ellos no interesa la forma funcional en sí, y se requiere tener la estimación de un patrón de comportamiento de los datos sobre el dominio X . Generalmente este tipo de aplicaciones cuenta con grandes volúmenes de observaciones sobre cambios pequeños en X .

Es interesante también cómo estos modelos cobran relevancia en el concepto de asociación para la toma de decisiones donde el ambiente de incertidumbre es alto. Se ha comprobado que la mayoría de las personas no suelen emplear reglas matemáticas para integrar la diversidad de variables en un momento específico, sino que se dejan llevar por estrategias intuitivas. Con frecuencia se hacen inferencias incorrectas en los juicios de asociación (Batanero et al., 2013).

Modelos para muestras complejas

Estos se desarrollaron debido a que los modelos lineales convencionales suelen asumir que la muestra de unidades de estudio tiene una estructura simple (aunque los modelos ANOVA y de efectos aleatorios consideran muestras estratificadas y por conglomerados). Bajo tales disposiciones, las muestras pueden tener un mayor nivel de complejidad, como el de anidamiento, o unidades de varios niveles (multinivel) o estructuras espacio-temporales o de estudios longitudinales. Entre ellos se cuenta al Modelo Lineal General Jerárquico (MLGJ) de Lindley y Smith (1972); y el modelo lineal general de efectos mixtos, el cual utiliza la fórmula $y_i = X_i W_i \Gamma + X_i u_i + e_i$ para realizar cálculos a través de matrices de varianzas y covarianzas (donde y_i es un vector de ni observaciones, e_i el correspondiente vector de errores aleatorios, X_i describe a la matriz de observaciones en las variables explicativas a nivel de unidad de estudio, W_i son las observaciones de las variables explicativas a nivel de conglomerados y u_i los correspondientes errores aleatorios).

Los modelos lineales generales de efectos mixtos han sido aplicados dentro del análisis de encuestas, estudios educativos, estudios de ambiente organizacional, estudios multinacionales y multicentro, estudios ecológicos, estudios de curvas de crecimiento y medidas repetidas y series de experimentos. Estos modelos tienen sus orígenes en estudios longitudinales y en análisis de series de tiempo, pero también se les ha vinculado con los modelos espaciales en geoestadística (Ojeda-Ramírez y Gutiérrez-Reyes, 2005).

Al respecto de estas aplicaciones, Lizasoain (2020) — en su estudio de los modelos estadísticos para la valoración de la eficacia escolar— utilizó una regresión multinivel empleando modelos jerárquicos lineales con base en cuatro criterios específicos: puntuaciones extremas, residuos extremos, crecimiento de puntuaciones y crecimiento de residuos. Ello facilitó tanto la evaluación como la comparación de diversos centros escolares a través de cuatro escalas de estimación distintas.

En esta categoría de modelos debe destacarse el uso del teorema de Bayes en los análisis estadísticos generales. En él los parámetros son las cantidades desconocidas de interés y por lo tanto se requiere especificar probabilidades alrededor de sus valores: indicadores del modelo, observaciones faltantes, mediciones inobservables directamente o imprecisas. De aquí se desprenden los modelos de hipótesis con base en la inferencia bayesiana.

Conviene aquí aclarar que el problema de elegir un modelo para describir el proceso de generación de los datos analizados es que este dependerá del problema y del propósito del estudio. En ciertas ocasiones, como resalta Gutiérrez Peña (2016), la forma en la que se obtuvieron los datos puede sugerir modelos apropiados como punto de partida (dicotómico Binomial, conteos Poisson, etc.). Sin embargo, frecuentemente el modelo debe reflejar una hipótesis que pueda ser verificada posteriormente en el contexto de los datos, constatando si las variables involucradas se interrelacionan linealmente o de otra forma.

...en el proceso de investigación

En el caso de la minería de datos, se han establecido dos fases para desarrollar un modelo estadístico de muestras complejas (Beltrán Martínez, 2015).

- 1a Fase. Comprobación de la precisión del modelo en datos independientes de los que serán utilizados.
- 2a Fase. Realización de una experiencia piloto con ese modelo, con el objeto de predecir la respuesta de los sujetos de estudio y de esta forma evaluar la fiabilidad de este modelo.

Este modelo puede requerir implementación en tiempo real, por lo que requiere ser descriptivo e interpretativo; además, al tener muchos usuarios debe difundirse y expresarse de manera comprensible para ser distribuido entre las muestras seleccionadas. Una vez resuelto este asunto, puede aplicarse la técnica denominada como *data mining*, proceso iterativo de extracción de información y patrones de comportamiento que permanecen ocultos entre grandes cantidades de información, a los que se les nombra descubrimientos (Beltrán Martínez, 2015).

Siguiendo con estos desarrollos, se han planteado generalizaciones que consideran simultáneamente las dimensiones espacial y temporal, considerando efectos fijos y aleatorios, pero también dando la posibilidad de considerar modelos paramétricos y no paramétricos. De hecho, ya se encuentran disponibles diversos programas informáticos (R, SPSS, S-PLUS, AMOS, STATA, Mplus, SAS, EQS, SAS Calis, entre otros), cuentan con procedimientos razonablemente sencillos para aplicar la mayoría de las técnicas y los modelos estadísticos, incluyendo con frecuencia al análisis multivariante de datos (Closas et al., 2013).

Análisis multivariante: Métodos descriptivos

La metodología estadística provee con otros procedimientos que contemplan el análisis de múltiples variables. Estos procedimientos se clasifican como descriptivos o de interdependencia debido a que las variables no están conceptualmente definidas como independientes o dependientes. Esencialmente se encargan de caracterizar o describir al objeto de estudio de forma conjunta a través de procedimientos numéricos. Usualmente, quién lleva a cabo la investigación tratará de recopilar la mayor cantidad de información acerca de su objeto de estudio; es decir, considerando el mayor número de variables posible. En este sentido, se requiere de procedimientos que logren capturar el comportamiento de las variables de forma conjunta.

Por otra parte, la investigación no requiere particularmente de la propuesta de un modelo estadístico cuando se tienen múltiples variables. Las técnicas multivariadas, al igual que los modelos estadísticos, tienen cierto grado de complejidad. Su aplicación práctica requiere del conocimiento de sus fundamentos y principios. En las técnicas de interdependencia, a diferencia de los modelos estadísticos, se considera que todas las variables asumen el mismo rol (no se hace distinción entre variables respuesta y explicativas). El fin es que a través de su aplicación se descubra cómo están interrelacionadas las variables y cuál es la estructura que subyace de la misma.

Su implementación y uso se ha extendido ampliamente, gracias al desarrollo del cómputo estadístico. Sin embargo, esto ha derivado en que quién

...en el proceso de investigación

investiga omite aspectos relevantes que preceden el análisis multivariante. Cada una de estas técnicas es aplicada de forma distinta y bajo diversas condiciones o circunstancias. No obstante, las fases en estos análisis requieren de atravesar por encrucijadas en las que el usuario debe tomar decisiones, las cuales deben sustentarse en el criterio estadístico. Pero esto, dependiendo del software que se desee usar, en ocasiones no pueden ser modificados o manipulados. En cierta medida se restringe el uso de pequeñas variaciones que mejoren el análisis. Por lo general los análisis exigen de la colaboración de alguien especialista en metodología estadística; por esta razón, para que la investigación aborde el camino del análisis multivariante, se recomienda recurrir al apoyo de la consulta estadística. Además, para utilizar estos procedimientos se requiere de una exploración asidua tanto de forma univariante como bivariante. Las técnicas de análisis multivariantes más habituales son las siguientes: análisis de componentes principales, análisis factorial (exploratorio y confirmatorio), análisis de correspondencia múltiple, análisis de agrupación (clúster) y el análisis de correlación canónica.

Análisis factorial (AF)

Constituye una de las técnicas más utilizadas para desarrollar y validar instrumentos de medición o test. Esta técnica permite esencialmente explorar de forma conjunta un grupo de variables. De esta forma logra identificar el número de factores comunes que pueden explicar las respuestas del instrumento y su composición.

Es decir, su aplicación permite identificar cuál es la estructura que subyace de la base de datos en relación a una serie de ítems. Algunos de los requisitos que deben cumplir los datos para su análisis por esta vía: (a) las variables deben ser cuantitativas de intervalo o de razón; (b) el número de observaciones debe ser por lo menos cinco veces el número de variables; (c) no se definen de antemano variables dependientes; y (d) debe existir correlación sin que se presente multicolinealidad. Como describe Gil Pascual (2015), este método es muy utilizado cuando se desea hacer una reducción de datos o cuando se persigue la búsqueda de un indicador (el coeficiente intelectual, el clima laboral, inteligencia emocional, la creatividad, etc.).

Antes de esto, se debe cerciorar de que los datos son susceptibles y útiles para ser analizados mediante esta técnica. Las variables deben estar relacionadas entre sí. Para esto, la matriz de correlaciones es sometida a distintas pruebas para su verificación mediante sus correlaciones, la prueba de esfericidad de Bartlett, su determinante y/o la prueba Kaiser Meyer-Olkin. Todas ellas son medidas que ayudan a determinar si las correlaciones parciales entre las variables son contundentes. Además, es necesario que el instrumento sea validado. Usualmente se utiliza el *alpha* de Cronbach. Una vez que la base de datos cumple satisfactoriamente estos requisitos, se puede aplicar el análisis factorial.

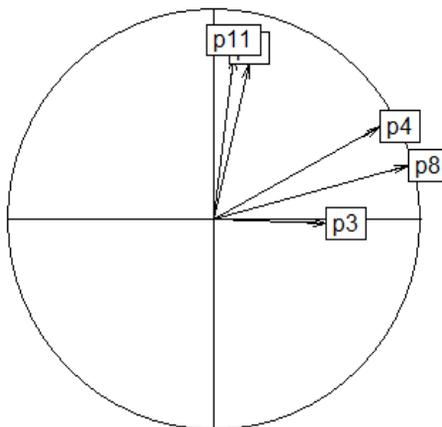
Para aplicarse, se debe partir del supuesto de que las variables, ítems o preguntas contribuyen a la formación de un determinado número de factores (o variables latentes). Es decir, se espera que cada una de las preguntas refleje en cierta medida una característica de

...en el proceso de investigación

un factor en particular. Se puede pensar en la aplicación de una encuesta a 300 estudiantes acerca de la salud mental. Su contenido serían 4 preguntas relacionadas con la ansiedad y otras 6 acerca de hábitos de estudio. Se esperaría que estos dos grupos de ítems recaigan en dos factores distintos. Cada uno de estos factores es ahora una variable latente con otros valores, los cuales albergan la información de las preguntas iniciales. De esta manera, se reduce el problema de 10 a únicamente 2 variables. Por ello, también es denominada como una técnica para la reducción de dimensionalidad; es decir, de variables. A continuación, se muestra un ejemplo sencillo pero completo de cuáles son los resultados a los que se llega y cómo pueden ser utilizados en el discurso de interpretación.

Se aplicó un cuestionario relacionado con la experiencia educativa de los estudiantes durante el periodo de la pandemia. Únicamente se realizaron 9 preguntas, de las cuáles se escogieron 5. Habiendo hecho las consideraciones mencionadas anteriormente, se confeccionaron los resultados mostrados en la Figura 7.

Figura 7. Despliegue de los vectores en el análisis factorial.



El par de variables o preguntas (p11 y p7) que se extienden sobre el eje vertical corresponden a preguntas que se relacionan con la percepción del apoyo de sus docentes para superar estas circunstancias especiales. Mientras que el segundo grupo de variables (p3, p4 y p8) que se extienden con mayor amplitud sobre el horizonte, son preguntas relacionadas con la satisfacción de su aprendizaje durante la pandemia. De esta forma, el análisis factorial revela de forma contundente cómo están estructuradas y relacionadas las variables en estudio. En este caso, la experiencia educativa de los estudiantes de educación superior durante la pandemia se puede explicar a través de dos factores: el “apoyo del docente” y la “satisfacción de su aprendizaje”. Estos factores también se denominan variables latentes. Se le llama variable latente a aquella que no puede ser observada directamente de la realidad, sino que se obtiene a partir de otras. El mismo estudio de estas nuevas variables puede desembocar en la elaboración de indicadores.

Análisis de componentes principales (ACP)

Otra de las técnicas encargadas de reducir la dimensionalidad es el ACP. A diferencia del anterior, dentro de este análisis es permitido usar la matriz de correlaciones o la de covarianzas. Los orígenes de este análisis parten de los ajustes ortogonales mediante mínimos cuadrados de Karl Pearson. Esta técnica sigue una metodología similar a la del análisis factorial. Además, busca encontrar un espacio de dimensión reducido capaz de representar la información de los

datos apropiadamente. El ACP transformará el conjunto de variables originales en nuevas variables latentes, de tal manera que estas últimas no estén correlacionadas. A éstas se les denomina componentes principales. Las nuevas variables son combinaciones lineales de las originales, de tal forma que recaban la máxima cantidad de información posible. Recordando que todas las variables deben ser numéricas, por lo general, los componentes principales se calculan a partir de las variables estandarizadas con media 0 y desviación estándar igual a 1. Si la base de datos la conforman variables en diferentes escalas, las variables con unidades más grandes tendrán una mayor influencia sobre las componentes calculadas; por esta razón, preferentemente se estandariza la base de datos.

Dentro de este análisis se debe decidir y considerar con cuántos componentes quedarse para explicar la variabilidad de los datos. Aunque en la mayoría de los casos se consideran solamente los dos o tres primeros componentes, que son los que albergan la mayor proporción de la varianza explicada. Esto permite representar los datos en dos o tres dimensiones. Evidentemente al pasar de, digamos, 16 variables a 2 componentes, hay una gran reducción de la dimensionalidad del problema lo cual posibilita representarlos de forma multivariante. Sin embargo, dentro de esta reducción existe una cierta pérdida en la variabilidad. En este sentido, la estrategia consiste en perder la menor cantidad de información posible.

Para calcular las componentes principales se utilizan las varianzas de cada variable. Esto provoca que el ACP sea sensible a los valores atípicos, por ello es necesario explorar los datos y anticipar este problema. Sin

embargo, esta revisión no se debe limitar a hacerla de forma unidimensional; hay que considerar más de dos variables. Por ejemplo, un sujeto que pesa 57 kilogramos no podría ser considerado como un dato anormal; lo mismo pasaría para un sujeto que mide 1.80 metros. Pero, si estas dos características pertenecen al mismo individuo, se estaría hablando de un sujeto atípico. Es por ello que los datos atípicos deben ser examinados de forma conjunta. Para realizar esto, la estadística provee con varias distancias que identifican a potenciales “outliers” de forma multivariante. La más popular es la de Mahalanobis. Esta medida es la distancia entre un punto o dato y la media que se ajusta en función de la correlación entre dimensiones.

Análisis de correspondencia múltiple (ACM)

Es un método multivariante que, al igual que el ACP y el AF, persigue el objetivo de explicar una gran cantidad de variables en un número de dimensiones reducido. La diferencia gira en torno a que el ACM requiere que las variables en la base de datos sean cualitativas y estén organizadas por categorías. Cuando únicamente se está hablando de dos variables categóricas, nos encontramos en el contexto de la correspondencia simple. En este caso, a través de tablas de contingencia, en donde las categorías de una variable aparecen como filas y las categorías de otra aparecen como columnas, el análisis resume la información de las categorías de ambas variables en un espacio, en el que se representan simultáneamente. Observando esta información en un nuevo plano se pueden comenzar a construir juicios

sobre cómo se corresponden las categorías de una variable con las categorías de la otra.

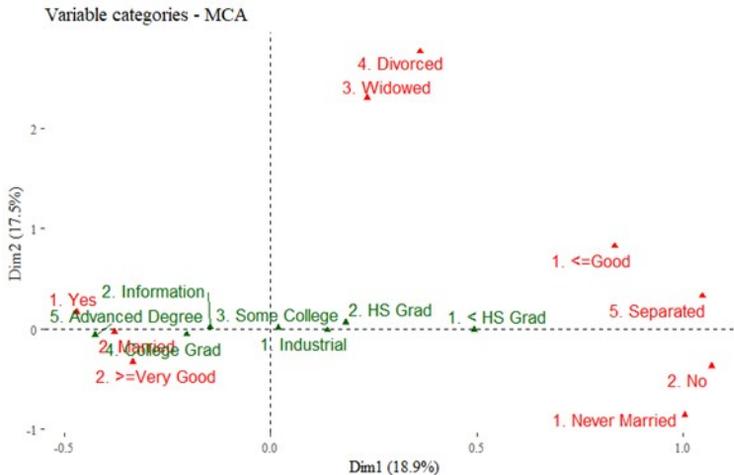
Esto mismo se extiende a tablas de contingencia multidimensionales o multivía, en los que se consideran 3 o más variables; en este escenario ya se está hablando de ACM. El ACM permite visualizar gráficamente las relaciones de dependencia e independencia de las categorías de un conjunto de variables. Para lograr determinar si existe una relación entre variables categóricas se utilizan contrastes de hipótesis de independencia; el que comúnmente se utiliza es la prueba Ji-cuadrada de Pearson. Es recomendable usar esta medida, ya que considera la ponderación de cada perfil con un peso. Es decir, cada fila o columna está ponderada con un peso que es proporcional a su relevancia dentro de un conjunto; evitando que se les otorgue mayor importancia a las categorías con pocas respuestas.

Como resultado final se presenta un plano bidimensional de coordenadas, en donde se pueden visualizar las categorías de todas las variables incluidas en el análisis, así como los puntos en donde se ubican los individuos u objetos de estudio. Su interpretación es intuitiva y muy útil para observar el comportamiento de forma conjunta. Aquellos puntos, o individuos que se encuentran más próximos tienen características similares. Por otra parte, cuando las categorías de distintas variables se encuentran próximas, indica que dichas características están presentes en conjunto recurrentemente por el grupo de individuos que se encuentre en ese espacio.

Este procedimiento estadístico es aplicado frecuentemente en los estudios de mercado, para

visualizar patrones de consumo, conocer las preferencias en compras de forma conjunta de clientes, así como identificar a aquellos consumidores que tienen hábitos similares, por ejemplo. Dentro del oficio de la investigación, el ACM se utiliza esencialmente para buscar tipologías que caractericen al objeto de estudio, siempre cuando se utilicen variables cualitativas o categóricas. Vale la pena recalcar que el proceso que conlleva su aplicación requiere del conocimiento de diversos aspectos como la relación de los perfiles marginales, columna y fila, la matriz de Burt y las contribuciones de cada categoría sobre la construcción de los ejes del plano bidimensional, por mencionar algunos. A continuación, se muestra un ejemplo, del producto final, al que aspira este análisis (Figura 8).

Figura 8. Análisis de correspondencia múltiple.



Nota: Elaboración propia.

...en el proceso de investigación

Como se puede observar, la Figura 8 muestra en dos dimensiones las categorías de distintas variables categóricas, en función de las respuestas o características de los individuos. Para el análisis se utilizó una base de datos que contiene la información de 3 000 trabajadores de la región del Atlántico medio. En este caso se quiere representar gráficamente la relación entre el nivel de estudio, el tipo de trabajo que tienen, su situación conyugal, su estado de salud y el estado de su seguro médico. Del lado izquierdo se aprecia cómo se aproximan las respuestas de que sí tienen seguro médico, un grado de estudios universitarios o mayor, salud mejor que buena, y que además trabajan en el área de la informática. Por otra parte, en la zona media aquellos trabajadores del sector industrial tienden a tener grados antecesores a los universitarios. Finalmente, se agrupa otro grupo cuya salud no es buena, y estos tienden a estar separados, a no haberse casado y a no tener seguro médico. Mediante este análisis, es claro que se podrían comenzar a construir tipologías y juicios acerca del colectivo de estudio, al ver cómo se relacionan las categorías de cada variable de forma conjunta.

Análisis de agrupación o clúster

Hoy en día la aplicación de los métodos de agrupación o clasificación se ha extendido. Antes de los avances computacionales llevar a cabo un análisis de este tipo resultaba extremadamente difícil, sobre todo cuando se trataban grandes conjuntos de datos. Esta técnica multivariante exploratoria consiste en la clasificación de los elementos de un colectivo de estudio en grupos,

considerando sus características. De tal forma que los elementos de un mismo grupo sean homogéneos, en la medida de la posible, pero heterogéneos con respecto a los elementos de otros grupos. Dicho de otra manera, el análisis tiene el propósito de reorganizar la información de los individuos en distintos grupos. Al hacer esto, se persigue el objetivo de acomodar las observaciones en grupos que no se conocen de antemano. La similitud entre objetos de un mismo grupo será plenamente alta, mientras que al comparar a objetos de diferentes grupos la similitud es baja.

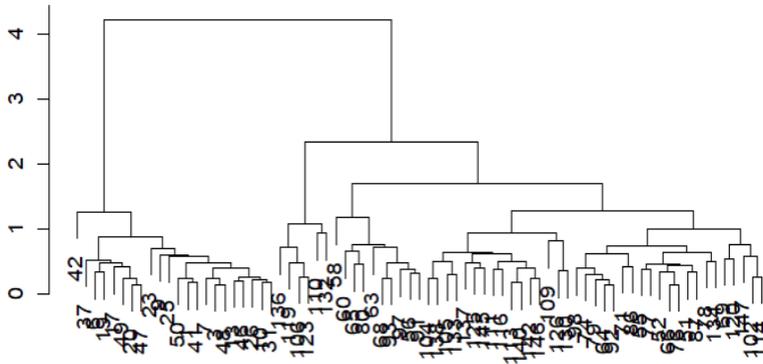
El análisis clúster cuenta con una amplia gama de métodos. Sin embargo, solamente son unos pocos los que se utilizan con mayor frecuencia. Esto deriva en que dichos métodos requieran un mayor estudio y experiencia en para su implementación. Aquí se distinguen dos distintas clases de métodos para hacer agrupaciones o clústeres: los métodos jerárquicos y no jerárquicos. Los métodos jerárquicos buscan crear grupos que posteriormente se irán agrupando en uno nuevo. A su vez estos métodos se subdividen en dos categorías: los aglomerativos y disociativos. Los primeros son considerados como ascendentes. Se van formando grupos de forma ascendente de tal forma que al final todas las observaciones y grupos formados se encuentran dentro de un solo grupo.

En cambio, los disociativos (o descendentes) parten de un grupo que contiene a todos los objetos individuales. A partir de ellos se van formando divisiones que dan origen a grupos que son cada vez más reducidos. La formación de clústeres o grupos en los métodos jerárquicos se basan en la matriz de distancias o similitudes. Estos métodos jerárquicos son útiles

...en el proceso de investigación

cuando se tiene el propósito de hacer agrupaciones de clústeres o para separar uno ya existente, el cual podría dar origen a otros dos clústeres (Figura 9).

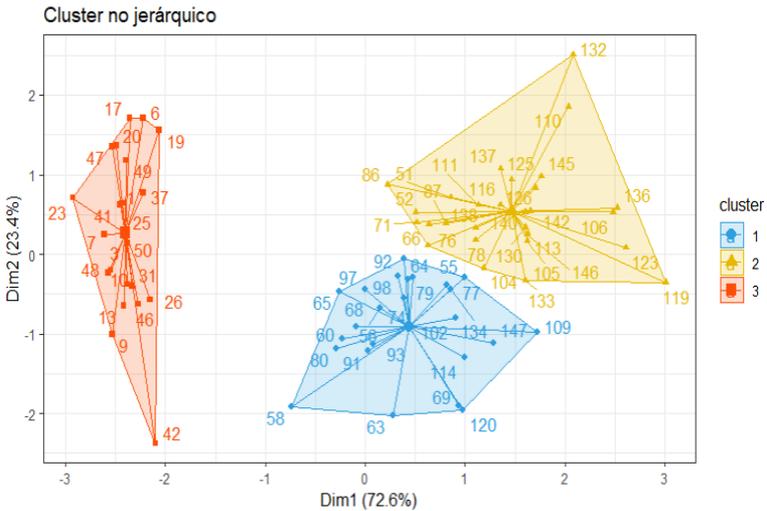
Figura 9. Clúster jerárquico.



Nota: Elaboración propia.

Por otra parte, los métodos no jerárquicos crean grupos completamente particionados. Una de las principales diferencias entre ambos métodos es que en los no jerárquicos la o el investigador deberá de indicar cuántos grupos son los que se deben de formar. Por lo tanto, para la formación de clústeres no jerárquicos requieren de un conocimiento previo. La agrupación de los elementos del colectivo se realiza partiendo de un número dado por quién lleva a cabo la investigación. Para ello, la metodología estadística provee con diversos recursos para escoger bajo un criterio el número óptimo de clústeres o grupos a formar. Además, estos métodos trabajan con la matriz de datos originales sin recurrir a la conversión de una matriz de distancias (Figura 10).

Figura 10. Clúster no jerárquico.



Nota: Elaboración propia.

Análisis de correlación canónica (ACC)

Al igual que los análisis de agrupación se han popularizado con la disponibilidad de programas estadísticos; derivando en un incremento de su aplicación en las investigaciones. Esta técnica multivariante se utiliza principalmente cuando se cuenta con diferentes variables dependientes. La correlación canónica permite determinar la relación entre un conjunto de variables cuantitativas predictivas sobre varias variables que se consideran como dependientes en el marco de la investigación. Dicho método fue acuñado por el estadístico Harold Hotelling alrededor de los años 40 del siglo XX. Hoy en día, constituye un análisis plenamente útil en áreas como biología, química,

psicometría, pedagogía y en sociología, por mencionar algunas. Esto ocurre cuando se busca analizar las relaciones entre múltiples variables, considerando que existe un grupo de dependientes y otro de independientes, previamente conceptualizadas.

Por otra parte, a través de otras medidas estadísticas permita cuantificar la relación entre dos conjuntos de variables. Dicho procedimiento cuenta con una serie de funciones que se denominan canónicas, las cuales se encargan de maximizar las correlaciones de las combinaciones lineales de ambos conjuntos. Estas funciones son las que determinan el grado de dependencia lineal entre los grupos de variables. Además, permiten dimensionar qué tan contundente es esta asociación. Es importante tomar en cuenta que las relaciones evaluadas por este análisis son lineales (aunque existen también métodos no lineales). Si la información en los datos está correlacionada de manera no lineal será necesario transformar las variables. Además, este análisis no requiere que las distribuciones de los datos cumplan el supuesto de normalidad. Sin embargo, se prefiere que sí la cumplan, debido a que con este atributo se tiene una mayor posibilidad de generalizar los resultados. En este sentido, al igual que muchos de los modelos estadísticos y técnicas multivariantes, el ACC cuenta con métodos para validar que los hallazgos encontrados pueden brindar información relevante acerca de la población y no sólo de los datos con los que se realizó el análisis. Para lograr esto es común que se recurra a la creación de dos submuestras extraídas de la base original. Se aplica el mismo análisis a cada una de estas nuevas muestras, y se comparan las funciones canónicas obtenidas y los

valores teóricos, entre otros aspectos relevantes. Con esta información es posible determinar si los resultados son similares o si difieren.

Particularmente, en este método se debe tomar ciertas precauciones al seleccionar el número de variables que habrá dentro de cada grupo. Como indican Badii et al. (2007), usualmente se ignora la influencia o el impacto que tiene incluir demasiadas variables en ambos grupos en relación al tamaño de la muestra. Por un lado, aquellas muestras que sean muy pequeñas no lograrán revelar las correlaciones apropiadamente, lo que puede desembocar en omitir la existencia de relación entre variables. Sin embargo, cuando se utilizan muestras grandes, el análisis tiende a indicar que existen correlaciones significativas, cuando probablemente sea un error. Cuando se desea aplicar la correlación canónica, quién realiza la investigación debe procurar que la división entre variables dependientes e independientes sea lo suficientemente clara; es decir, que, desde el enfoque teórico del problema, se asume que probablemente exista una relación entre ambos grupos.

Conclusiones

La modelación estadística y el análisis multivariante constituyen dos pilares en el área de investigación. En este campo el conocimiento experto de diversos fenómenos estudiados juega un papel determinante. De tal modo, mediante el enfoque multidisciplinario, se promueve el desarrollo de esta perspectiva de la metodología estadística, como asevera Ojeda (2004). Algunas de sus áreas de mayor desarrollo son la minería

de datos y la teoría de las decisiones a partir de la inferencia Bayesiana. Ambas son ejes que empujan proyectos y nuevas contribuciones al análisis de variables numerosas, complejas e interrelacionadas. La aplicación de las herramientas y procedimientos estadísticos pueden resultar bastante efectivas y útiles para resolver problemas con mayores grados de complejidad.

Es pertinente hacer énfasis en que quienes desempeñan el oficio de la investigación se apropien de la amplia gama de recursos que ofrece la metodología estadística, con la esperanza de que sus beneficios y utilidades se amplifiquen y sean aprovechadas en nombre del avance científico. Implementar los procedimientos estadísticos se ha convertido en una actividad que requiere de un menor grado de esfuerzo, gracias al avance del cómputo estadístico. Sin embargo, esta facilidad en su acceso ha desembocado en el desinterés por comprender ciertos aspectos que, al no ser considerados apropiadamente, repercuten con gran impacto en los resultados finales. Por ello, su aplicación exige de cierto grado de conocimiento sobre sus fundamentos, evitando caer en errores y en resultados no verdaderos, que poco contribuirán ciclo de la investigación y que terminarán por descalificar el uso de la metodología estadística.

De forma específica, la importancia de los modelos estadísticos se destaca por ayudar a fomentar el razonamiento deductivo (en el caso de la aplicación de las herramientas estadísticas inferenciales) e inductivo (para el caso de diseños y pruebas de hipótesis). El objetivo es identificar las fuentes que producen variaciones (unidades de medida, tipos de datos,

técnicas de muestreo, condicionantes y criterios de análisis, entre otros) entre los grupos o fenómenos a comparar; estas variaciones pueden ser producidas por factores generales (limitación de instrumentos de recolección de datos, particularidades del lenguaje, etc.) o específicas (características de los sujetos, circunstancias contextuales o aspectos no considerados).

Desprendiéndose de todo lo anterior, como lo señalan Batanero et al. (2013), la consolidación del razonamiento estadístico “permite buscar explicaciones y causas para la variación y realizar inferencias y predicciones, con un cierto margen de error, teniendo en cuenta la variación no explicada o aleatoria”. Consideraciones que deben ser tomadas en cuenta al momento de constituir los modelos estadísticos como “representaciones de la realidad”. Además de diseñar instrumentos para comprender al colectivo o al fenómeno de estudio, sabiendo diferenciar las condiciones del modelo de la complejidad de los hechos sociales y educativos. En este sentido, usar apropiadamente la estadística permite comprender el comportamiento de fenómenos difíciles de estudiar.

Por último, la capacidad de integrar todos los datos en un modelo estadístico es un componente esencial a partir de las fases iniciales (planteamiento del modelo) hasta las etapas finales (interpretación del modelo en la realidad) del ciclo de modelización. Como consecuencia directa de ello, se coincide con Harris (2021) en cuanto a que los especialistas en CS y ciencias de la educación deben aprovechar las ventajas de los métodos mixtos condensados en modelos estadísticos para estudiar temáticas complejas. Esto puede ayudar a valorar los

...en el proceso de investigación

efectos de los factores culturales y los desafíos prácticos de los países en contextos en desarrollo, como es el caso de México.

Con base a lo antes expuesto, la elección y las combinaciones de métodos y modelos deben guiarse por la integración funcional de las preguntas sobre el problema de estudio, los objetivos y las hipótesis. Sin duda no es una tarea fácil, ya que quienes investigan deben aprender, reflexionar y actualizarse en el manejo de las TIC que les ayuden a manejar mayor cantidad de datos y de mejor manera. También es indispensable que ellos incorporen ciclos de retroalimentación. Con ello se busca conducir a cambios profundos en los planteamientos iniciales de su investigación, retándolos a entremezclar sus resultados cualitativos con los cuantitativos a través de una secuencia metodológica dinámica, creativa y en constante evaluación.

Referencias

Mohammad Badii y Jaime Arturo Castillo, “Análisis de correlación canónica (ACC) e investigación científica”, *InnOvaciOnes de NegOciOs*, núm. 8, julio-diciembre 2007, pp. 405-422.

<http://eprints.uanl.mx/12486/1/A9%20%281%29.pdf>

Adán Barreto-Villanueva, El progreso de la estadística y su utilidad en la evaluación del desarrollo, *Papeles de población*, núm. 73, julio-septiembre de 2012, pp. 241-271.

<http://www.scielo.org.mx/pdf/pp/v18n73/v18n73a10.pdf>

- Carmen Batanero, Carmen Díaz, José Miguel Contreras, y Rafael Roa, "El sentido estadístico y su desarrollo" *Números, Revista de didáctica de las Matemáticas*, núm.1, julio de 2013, pp. 7-18.
<http://funes.uniandes.edu.co/3651/1/Batanero2013EINumeros83.pdf>
- Beatriz Beltrán Martínez, *Minería de datos* [Tesis de maestría en Sistemas Computacionales, Benemérita Universidad Autónoma de Puebla], 2015. Consultado en:
<http://bbeltran.cs.buap.mx/NotasMD.pdf>
- María Castro y Luis Lizasoain, "Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales", *Revista Española de Pedagogía*, núm. 251, enero-abril de 2012, pp. 131-148.
- Menacho Chiok y Cesar Higinio, "Modelos de regresión lineal con redes neuronales", *Anales Científicos de la Universidad Nacional Agraria La Molina*, núm. 2, julio-diciembre de 2014, pp. 253-260.
<https://dialnet.unirioja.es/servlet/articulo?codigo=7025156>
- Antonio Humberto Closas, Edgardo Alberto Arriola, Cristina Isabel Kuc, Mariela Rosana Amarilla y Ethel Carina Jovanovich, "Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría", *Enfoques*, núm. 1, septiembre-diciembre de 2013 pp. 65-92.
<http://www.scielo.org.ar/pdf/enfoques/v25n1/v25n1a05.pdf>
- Juan Del Valle Moreno y Walkiria Guerra Bustillo, "La multicolinealidad en modelos de regresión lineal

...en el proceso de investigación

múltiple”, *Revista Ciencias Técnicas Agropecuarias*, núm. 4, octubre-diciembre de 2012, pp. 80-83.

<http://scielo.sld.cu/pdf/rcta/v21n4/rcta13412.pdf>

Juan Antonio Gil Pascual, *Metodología cuantitativa en educación*, España, UNED, 2015.

Jamelia Harris, “Mixed methods research in developing country contexts: Lessons from field research in six countries across Africa and The Caribbean”, *Mixed Methods Research*, núm. 2, julio de 2021, pp. 165-182.

<https://doi.org/10.1177/15586898211032825>

Luis Lizasoain, “Criterios y modelos estadísticos de eficacia escolar”, *Revista de Investigación Educativa*, núm. 2, 2020, pp. 311-327.

<https://revistas.um.es/rie/article/view/417881>

Yoonkyung Lee & John Nelder, “Hierarchical generalized linear models”, *Journal of the Royal Statistical Society*, núm. 58 B, 1996, pp. 619-678.

Dennis Lindley & Adrian Smith, “Bayes estimates for the linear model”, *Journal of the Royal Statistical Society*, núm. 34 B, 1972, pp. 1-41.

Modelos de regresión lineal múltiple. (Roberto Montero Granados, 2016). Consultado en:

http://www.ugr.es/~montero/matematicas/regresion_lineal.pdf

John Nelder & Robert Wedderburn, “Generalized linear models”, *Journal of the Royal Statistical Society*, núm. 3, 1975, pp. 370-384.

<https://www.jstor.org/stable/2344614>

La modelación estadística (Mario Miguel Ojeda Ramírez, 2004). Universidad Juárez Autónoma de Tabasco. Consultado en:

<https://www.researchgate.net/publication/295010680>

Mario Miguel Ojeda-Ramírez y Gildardo Gutiérrez-Reyes, “Propuesta de un modelo lineal jerárquico para analizar una serie de experimentos”, *Agrociencia*, núm. 39, agosto de 2005, pp. 529-538.

Juan Rositas Martínez, “Los tamaños de las muestras en encuestas de las ciencias sociales y su repercusión en la generación del conocimiento”, *Innovaciones de negocios*, núm. 22, noviembre 2014, pp. 235-268. <http://eprints.uanl.mx/12605/>

Paul Schaffer, *Q-Squared: Combinación de enfoques cualitativos y cuantitativos en el análisis de la pobreza*, Reino Unido, Oxford University Press, 2013.

<https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199676903.001.0001/acprof-9780199676903>

Interpretación de resultados y elaboración de conclusiones

LEOMAR MAR MEDINA

Introducción

La preparación de la base de datos y la aplicación del análisis estadístico de forma apropiada llevará a realizar la última fase: interpretar los resultados y construir conclusiones que permitan cerrar de manera contundente con la investigación. Previamente se habló de la diversidad de análisis que se pueden implementar para extraer conocimiento del colectivo de estudio, según sea el caso. Estos análisis producirán ciertas salidas o resultados estadísticos, los cuales deberán ser traducidos y leídos desde el marco referencial de la investigación. En el presente capítulo se abordará la importancia de realizar la interpretación de resultados dentro del proceso de investigación, así como la formulación de manera concisa de las conclusiones.

Algunos preliminares

La interpretación de los hallazgos y los resultados obtenidos debe ser concisa. El fin es que la transmisión de los mismos se convierta en una actividad más digerible y eficaz para aquellos que lean lo redactado. La buena interpretación y la contundencia de las

conclusiones depende en gran medida de lo que se ha hecho en todas las etapas por las que atravesó el estudio. Es decir, la selección del diseño estadístico de la investigación, la selección de la muestra o colectivo de estudio, la recolección de los datos y otras evidencias, su tratamiento y preparación, hasta llegar a la implementación de la estrategia de análisis. El desarrollo del discurso conclusivo debe construirse contemplando cómo se llevó a cabo la implementación de la metodología estadística, pero así mismo las preguntas de investigación, los objetivos y la hipótesis de trabajo. Además de todo, será fundamental la comprensión del problema, su delimitación y la sustentación en un marco teórico, así como la revisión de los conocimientos previos disponibles sobre el tema que se está abordando.

Cuando ya se han obtenido los resultados de los análisis estadísticos se requiere tratarlos adecuadamente para obtener, una síntesis de las consecuencias más importantes. En este sentido, es preciso llevar a cabo una interpretación clara y concisa de los hallazgos encontrados a la luz de las preguntas de investigación. El buen manejo e interpretación de los resultados funciona como la directriz para construir las conclusiones del estudio. Evidentemente, quién lleva a cabo la investigación tiene la intención de que sus resultados sean generalizables sobre la población objetivo y no solamente en el caso específico de su muestra o los casos que se usaron para obtener los datos. La generalización de los resultados (validez externa) está determinada por cómo se llevaron los procedimientos estadísticos y sus circunstancias (a esto se le denomina la validez interna). Por lo tanto, la elaboración de las conclusiones requiere de comprender profundamente el

problema, el marco referencial, y también la metodología que se utilizó.

Los propósitos de las investigaciones varían en función del contexto en el que se estén aplicando. Sin embargo, en absoluto, todos persiguen un objetivo común: el de obtener más conocimiento acerca de un colectivo de interés. Ojeda y Behar (2006) describen importantes aspectos al momento de juzgar la validez de un proceso por el cual se generó conocimiento. Por ejemplo, la calidad de los datos obtenidos. Si los datos no fueron obtenidos con cierta confiabilidad de que efectivamente están midiendo los elementos de la realidad que se desea abordar, es de esperarse que el estudio termine por dirigir a conclusiones que no son necesariamente válidas. Cuando esta materia prima es registrada confiablemente, se dice que el estudio cuenta con validez externa. Por otra parte, generar nuevas afirmaciones requiere de tener cierta compatibilidad con la teoría expuesta, o lo que hasta el momento se conoce como verdadero. En este contexto, si se encuentran contradicciones entre los resultados y la teoría, es necesario hacer una valoración para rechazar las nuevas afirmaciones (buscando razones), o se replantea la teoría aceptada hasta el momento, explicando el nuevo comportamiento. A esto se le denomina validez interna; esto es cuando a través de la lógica de la inferencia estadística se hace una comparación entre lo preestablecido y lo encontrado.

El objetivo de la interpretación de los resultados es darle un sentido y buscar un significado concreto a las respuestas, asociándolo a otros conocimientos disponibles recapitulando las fases que preceden (Alva Santos, s.f.). Los resultados obtenidos de los análisis se

confrontan con las hipótesis formuladas considerando al mismo tiempo los procedimientos de la investigación. Las conclusiones deben estar conectadas directamente con las preguntas que dieron origen a la investigación. Al momento de desarrollar las conclusiones, quien investiga necesariamente debe contemplar factores externos que puedan influir sobre los resultados obtenidos.

Como describe Bernal (2010), el análisis de resultados consiste en interpretar los hallazgos relacionados con el problema de investigación, los objetivos establecidos, la hipótesis y/o preguntas formuladas, y las teorías o presupuestos planteados en el marco teórico. La finalidad es evaluar si se respalda lo presentado en las teorías o no, o si se generan debates con la teoría que ya se ha establecido. La interpretación de los resultados nos permite encontrar fortalezas que coadyuven en la búsqueda de soluciones a las diversas problemáticas estudiadas. Lo anterior mediante nuevas vetas de conocimiento que dan pauta para la creación de nuevas aportaciones en cada área. De igual manera, Bernal (2010) expresa que al analizar los resultados obtenidos debe indicarse si el estudio respondió o no a las hipótesis o preguntas planteadas para desarrollar los objetivos del estudio. En caso de no encontrar sustento en la hipótesis, cabe la posibilidad de que se ha creado una nueva veta. No quiere decir que está mal nuestro estudio, sino que tal vez es momento de llevar a cabo algunas adecuaciones.

Para las investigaciones cuyos resultados son de una aplicación de la metodología estadística, el conocimiento de los fundamentos estadísticos se convierte en un recurso indispensable para traducirlos a un lenguaje no técnico y presentarlos de manera

correcta. Para cumplir con este propósito, es recomendable que quien investiga se apoye de recursos como cuadros, tablas y gráficos. Debe procurar seleccionar sólo los elementos más importantes o necesarios para explicar los nuevos hallazgos.

De igual manera Taylor y Bogdan (1986, como se citó en Vilchez, 2007), plantean que el tratamiento de los datos y de la información se debe llevar a través de un análisis comprensivo. Construyendo un discurso cimentado con la comprensión y rastreo de la información, mediante la búsqueda de categorías fundamentales en los hechos que se han descrito. Para la interpretación de los resultados se debe utilizar un lenguaje técnico pero traducible a uno no técnico para lograr que los lectores entiendan y comprendan lo que se les desea transmitir. Así mismo, el apoyo de los gráficos adecuados ayudará para que la información no se desvirtúe y se cumpla en transmitir el conocimiento deseado.

Interpretación de resultados

La interpretación de resultados es considerada la última etapa de la tarea investigativa, siendo la antesala que detalla y da pie a las conclusiones. Este apartado es donde se muestran los hallazgos obtenidos, por lo que es necesario hacerlos hablar. Las conclusiones son la esencia principal de la interpretación de resultados (Alva Santos, s. f.). Consiste en dar sentido a la información obtenida, tratándola y organizándola para poder explicar, predecir e interpretar el fenómeno objeto de estudio y proponer una serie de respuestas al problema planteado (Mosteiro García y Porto Castro, 2017). Por consiguiente,

este apartado nos ayuda a dar el sentido de la investigación y a contrastar empíricamente las hipótesis de trabajo, los objetivos y argumentar con respuesta a la pregunta de investigación.

De acuerdo con Rojas-Soriano (2013, como se citó en Gallardo, 2017), la interpretación de resultados es un proceso ligado al análisis de los datos, pero “consiste en separar los elementos básicos de la información y examinarlos con el propósito de responder a las distintas cuestiones planteadas en la investigación”. La interpretación exige de un esfuerzo intelectual, en el cual, a la información empírica recabada, se le otorga un significado detallado.

Analizar los datos tiene como fin último llevar a cabo un extracto de lo observado durante nuestra tarea investigativa. Tiene como alcance el dar respuestas a todas aquellas interrogantes planteadas al inicio de la investigación. De acuerdo con Sellitz (1970, como se citó en Alva Santos, s.f.) se debe buscar un significado más amplio a las respuestas mediante el trabajo con otros conocimientos disponibles. Esto es la base para definir y clarificar los conceptos y las relaciones entre variables, así como los hechos que son materia de la investigación.

La forma que adquiere el análisis e interpretación de los resultados varían de un estudio a otro, de forma aislada y en conjunto. Lo anterior depende de los distintos esquemas o niveles de investigación y, fundamentalmente, del diseño propuesto. No se debe olvidar que, los datos, a partir de los cuales se inicia el análisis, son diferentes según el nivel de elaboración realizado. Dependen de la naturaleza del problema de investigación, del tipo de investigación, así como de las

técnicas y procedimientos utilizados en la elaboración (Alva Santos, s.f.).

Para Palella y Martins (2012) la interpretación de los resultados consiste en hacer inferencias acerca de los datos ya procesados; examinando cómo se integra lo observado dentro de un contexto teórico. Para estos mismos autores la interpretación depende de tres factores: (a) el nivel de medición de las variables; (b) la manera como se hayan formulado las hipótesis; y (c) el interés de quien investiga.

Al tratar de interpretar los resultados es indispensable acudir a la teoría suscitada en el planteamiento del problema y en su marco conceptual. De igual manera, se revisan los objetivos que se plantearon, y se establecen comparaciones o asociaciones de los resultados obtenidos con otros en condiciones similares, o se contrasta lo encontrado en la revisión bibliográfica del conocimiento previo (Palella y Martins, 2012). “Al interpretar los resultados es necesario tener en cuenta que se debe ofrecer una discusión para cada uno de los objetivos y/o hipótesis planteados”. También, debe expresar “en qué medida los resultados avalan o no los objetivos o hipótesis planteadas” (Palella y Martins, 2012, p. 183).

Interpretación de análisis cualitativos

Algunos autores (Izcara Palacios, 2014; Sierra-Bravo, 2007) señalan que las investigaciones que se abordan desde en el enfoque cuantitativo, la interpretación suele ser más objetiva pues los resultados son cuantificables. Por otra parte, el enfoque cualitativo puede estar sujeto adquirir un carácter más subjetivo. En este último, la

capacidad de intuición personal, la comprensión del problema y el conocimiento en el área desempeñan un papel clave para elaborar interpretaciones enriquecedoras. Por ello, interpretar los resultados del análisis cualitativo implica contemplar las características del diseño de la investigación y cómo se desarrollo en la práctica (Chárriez, 2012).

El sesgo de quien investiga es uno de los aspectos de la investigación con énfasis cualitativo que ha llevado a la idea de que esta es más bien subjetiva y no tanto objetiva (Kawulich, 2005); en su afán de capturar la esencia de los fenómenos sociales inmediatamente, puede perder la distancia y ofrecer interpretaciones excesivamente personales (Izcara Palacios, 2014; Ruiz-Olabuénaga, 2003). Cuando la investigación es abordada desde el enfoque cualitativo, la perspectiva humanista está implícitamente integrada. Pues, se trata de comprender la conducta humana, según los fundamentos de la fenomenología, el existencialismo y la hermenéutica (Chárriez, 2012).

Sobre las conclusiones

En primer lugar, las conclusiones no son una recapitulación de los resultados; éstos ya se presentaron e interpretaron. En términos simples, el apartado de conclusiones recoge una interpretación no técnica de los resultados obtenidos (Mosteiro García y Porto Castro, 2017). Además, se hacen generalizaciones, se sintetiza el conocimiento obtenido y se habla acerca de las implicaciones prácticas; por consiguiente, en este apartado quién investiga explica el significado que tienen los resultados con base en los objetivos y la pregunta de

...en el proceso de investigación

investigación; es aquí donde podemos incluir algunas líneas para realizar investigación a futuro.

Las conclusiones se elaboran específicamente en función del problema de investigación, los objetivos particulares y las hipótesis planteadas, cuando las hay. Estas conclusiones deben destacar los aspectos más relevantes de los resultados del trabajo realizado, explicando su relación con los contenidos del marco teórico. De acuerdo con Pineda Beatriz et al. (1994), las conclusiones y recomendaciones constituyen la presentación de los hallazgos y sugerencias más sobresalientes de la investigación. Preferentemente las conclusiones deben expresarse de forma concreta y sencilla, sin olvidar mencionar cómo se relacionan con el problema y los objetivos planteados al inicio de la investigación.

En conclusión

Cuando la metodología estadística es aplicada apropiadamente, sus virtudes se verán reflejadas en la interpretación de los resultados. Esto debe llevarse a cabo con sumo cuidado, ya que es aquí de donde se desemboca la toma de decisiones y las recomendaciones prácticas para intervenir, de alguna forma, en la realidad. Indudablemente, es esta la parte que proporciona el conocimiento que se requería cuando se diseñó el estudio. Es importante destacar que la lectura de los resultados requiere de la revisión detallada del marco referencial; es decir, las teorías y el conocimiento de investigaciones previas sobre el tema. Contrastar adecuadamente el estado del conocimiento actual con las evidencias empíricas encontradas, deriva

en que el conocimiento respecto al tema se enriquezca. Por esta razón se dice que los resultados se deben discutir para converger a buenas conclusiones. Es vital que este contraste entre el conocimiento disponible y los resultados encontrados se haga de forma minuciosa, ya que es en este punto posible detectar nuevas oportunidades de investigación, generando así directrices de mayor profundidad para proseguir con el ciclo investigativo. No se debe tener temor a encontrar desacuerdos entre los resultados y el estado del conocimiento. La explicación e interpretación de las diferencias, en ocasiones, da un sentido más profundo y rico a la investigación. De la misma forma, es fundamental no olvidar que quien investiga debe estar dotado de recursos y elementos que mejor le permitan presentar estos resultados, tanto de forma visual como escrita. En conclusión, la “forma” y el “mensaje” deben estar en complicidad. Llevar esto apropiadamente conllevará a cerrar la investigación correctamente.

Referencias

- Análisis de los datos e interpretación de resultados*
(Antonio Alva Santos). Consultado en:
https://cmapspublic2.ihmc.us/rid=11772768992171477413697_5143/analisdatosinterpretac-1.pdf
- Cesar Bernal, *Metodología de investigación. Administración, economía, humanidades y ciencias sociales*, Colombia, Pearson Educación, 2010.
- Mayra Chárriez Cordero, “Historias de vida: Una metodología de investigación cualitativa”, *Revista Griot*, núm. 1, diciembre de 2012, pp. 50-67.

...en el proceso de investigación

<https://revistas.upr.edu/index.php/griot/article/view/1775>

Esther Gallardo, *Metodología de la investigación: manual autoinformativo interactivo*, Perú, Universidad Continental, 2017.

Simón Pedro Izcara Palacios, *Manual de investigación cualitativa*, México, Fontamara, 2014.

Barbara Kawulich, “La observación participante como método de recolección de datos”, *Forum Qualitative Social Research*, núm. 2, mayo de 2005, pp. 1-28. <https://www.qualitative-research.net/index.php/fqs/article/view/466/997>

María Josefa Mosteiro García y Ana María Porto Castro, “La investigación en educación”, en *Notas teórico-metodológicas de investigación en educación: conceptos y trayectorias*, Leila Pio Mororó, María Elizabete Souza y Raimunda Moreira de Assis (orgs.), Brasil, Editora DA UESC, 2017, pp. 13-40.

Santa Palella y Feliberto Martins, *Metodología de la investigación cuantitativa*, Venezuela, Fondo Editorial de la Universidad Pedagógica Experimental Libertador, 2016.

Mario Miguel Ojeda Ramírez y Roberto Behar, *Estadística, productividad y calidad*, México, Secretaría de Educación de Veracruz, 2006.

Elia Pineda Beatriz, Eva Luz de Alvarado y Francisca de Canales, *Metodología de la investigación. Manual para el desarrollo de personal de salud*, Estados Unidos, Organización Panamericana de la Salud, 1994.

José Ignacio Ruiz-Olabuénaga, *Metodología de la investigación cualitativa*, España, Universidad de Deusto Bilbao, 2003.

Restituo Sierra-Bravo, *Técnicas de investigación social: Teoría y ejercicios*, España, International Thomson Editores y Paraninfo, S.A, 2007.

Nieves Vilchez, *Enseñanza de la geometría con utilización de recursos multimedia. Aplicación a la primera etapa de educación básica* [Tesis de doctorado, Universitat Rovira I Virgili]. Repositorio institucional, 2004. Consultado en:

<https://www.tdx.cat/handle/10803/8928#page=1>

Escritura del reporte y unificación editorial

VÍCTOR HUGO RAMÍREZ RAMÍREZ
MARIO MIGUEL OJEDA RAMÍREZ

Introducción

La redacción del reporte representa el punto culminante de todo el proceso de investigación. Significa que la investigación documental está concluida, que se aplicaron las técnicas de investigación de campo apropiadas y se cuenta con los suficientes datos para llevar a cabo el análisis estadístico adecuado. También que este se ha realizado y se tienen los resultados suficientes, así como su interpretación y discusión, la cual derivó en conclusiones, recomendaciones y establecimiento de líneas futuras y limitaciones del estudio. Como todo proceso de creación, la escritura no es un proceso lineal. De entrada, cuando se desarrolla el protocolo es posible contar con ciertas secciones que presentan, si bien de manera esquemática, ciertas informaciones que se pueden llevar hasta el reporte. Puede ser el caso de la fundamentación teórica y la propuesta metodológica, sobre todo, pero regularmente sólo se toma como un buen antecedente. Es así que el reporte final debe, sin embargo, cumplir con una serie de elementos a tomar en cuenta desde el inicio mismo de la

investigación. Asimismo, tiene que recuperar la experiencia y resultados durante el proceso, y realizar un esfuerzo especial de sistematización y síntesis; todo esto tiene por fin evitar omisiones al momento de reunir toda la evidencia requerida en la redacción de este documento, que es la culminación del proceso investigativo. En este capítulo se dará un repaso de esta fase, sus conceptos y algunas recomendaciones para realizarla de manera correcta.

Consideraciones previas a la redacción

Un reporte de investigación puede definirse como un documento escrito de variada extensión y con una estructura determinada. Su objetivo es hacer una contribución al conocimiento de un tema mediante la presentación de los resultados de un proceso de investigación. Los formatos para el reporte de investigación varían de acuerdo con su etapa, contexto y propósito de difusión. Por este motivo, no se puede hablar de la existencia de un formato único, sino más bien de recomendaciones (Ojeda Ramírez et al., 2011). En lo referente a la etapa, los reportes de investigación se pueden clasificar en preliminares, parciales o finales (Acosta León, 2012). En el primer caso no se trata de presentar resultados definitivos, sino avances generalmente por petición de alguna parte interesada; en el segundo, un estudio puede abarcar diferentes aspectos, pero sólo se requiere presentar uno de ellos y no la totalidad de los resultados (por ejemplo, aspectos socioeconómicos en un estudio poblacional); finalmente, en el tercer caso, se incluye información detallada de los

diversos componentes de la investigación, así como de los resultados.

Sobre los contextos, generalmente se reconoce que los reportes se dividen entre los de contexto académico y no académico. El segundo tipo se refiere principalmente a situaciones de tipo comercial (como puede ser un estudio sobre un producto o un servicio) o un asunto para difusión entre el público en general que requiera el uso de lenguaje comprensible para todos (por ejemplo, la importancia de la esterilización de mascotas). En el contexto académico, sin embargo, el reporte se presenta a lectores que conocen o están inmersos en el tema; esto generalmente sucede entre los miembros de una comunidad académica de una institución de educación superior (estudiantes, profesores, investigadoras, investigadores, autoridades). Este último punto determina también el formato dentro del cual se presenta el reporte, pues, si bien los destinatarios pueden ser similares, el propósito de difusión puede variar. Algunos de estos propósitos pueden ser los siguientes:

- Tesis de grado y posgrado. Es el documento que se presenta al concluir un ciclo de estudios en nivel de licenciatura, maestría o doctorado. Como en todo documento de naturaleza académica, los formatos y lineamientos a seguir son rigurosos y cada institución de educación superior determina las reglas y requisitos que dicho documento debe cumplir. La extensión varía enormemente entre instituciones, disciplinas y grados.
- Artículo de revista científica o capítulo de libro. En este caso el documento aborda un tema específico relacionado con la publicación. En ocasiones, este tipo de publicaciones se derivan de investigaciones más amplias como las que se llevan a cabo para la

realización de una tesis de grado. El formato y los lineamientos de redacción están generalmente dictaminados por quien edita el libro o revista.

- Ponencias o conferencias. Se trata de una situación muy similar a la anterior. La diferencia es que la publicación se acompaña de una presentación en vivo (presencial o virtual), donde se presentan los puntos más relevantes del reporte. Quienes organizan el evento, llámese congreso, simposio, foro o encuentro, entre otros, establecen los lineamientos de publicación. A diferencia de las revistas científicas, los resultados de estos eventos de tipo académico generalmente dan a conocer estas aportaciones en documentos llamados memorias o actas.
- Reporte de investigación aplicada. Generalmente este tipo de reportes tratan sobre temas de interés para una sociedad o un grupo específico, el cual solicita a un especialista o grupo de especialistas su elaboración. A menudo el fin de esta investigación es llegar a una toma de decisiones informada basada en la evidencia presentada en el documento.
- Reporte para curso escolar. En programas de grado y posgrado es común encontrar materias donde se requiere la elaboración de un reporte de investigación como parte de un curso de metodología. Ejemplos de este tipo son los que proveen Gómez Chagoya (2004) para la Universidad Autónoma del Estado de México (UAEM) y Ojeda Ramírez et al. (2011) para la Universidad Veracruzana.

El formato que se presenta aquí es de un reporte académico final que formó parte del Seminario de Estudios del Doctorado en Innovación en Educación Superior que ofrece el Centro de Investigación e Innovación en Educación Superior de la Universidad Veracruzana, a cargo de los profesores Mario Miguel

...en el proceso de investigación

Ojeda y Juan Carlos Ortega. Por tanto, es una creación contextualizada en las necesidades específicas de esta experiencia educativa.

Los elementos constitutivos del reporte de investigación

Como se puede intuir a partir de la sección anterior, los elementos constitutivos del reporte de investigación varían grandemente. No sólo entre disciplinas e instituciones o por los propósitos de difusión, sino que dichas variaciones pueden existir de una materia a otra o dentro de un mismo programa educativo, inclusive. A continuación, se presentan los elementos que aparecen de forma recurrente de acuerdo con las propuestas de diversos autores (Acosta León, 2012; Arias, 2012; Gómez Chagoya, 2004; Hernández Sampieri et al., 2014; Tena Suck y Turnbull Plaza, 2007). Dentro de cada apartado se explica de manera sucinta en qué consiste cada uno de ellos.

1. Aspectos preliminares

Son aquellos que dan información general acerca del contenido del reporte antes de adentrarse en el contenido principal. Consisten principalmente en la portada, índice e introducción. En ocasiones se puede incluir un prólogo.

1.1 Portada

Los elementos que deben estar presentes en la portada son los siguientes: el nombre de la institución y

la entidad o dependencia específica que alberga el programa dentro del cual se lleva a cabo la investigación; título del reporte; autor(a) o autores (as), asesores(as), etc.; nombre de quien es responsable del curso o tutor(a); y lugar y fecha.

En algunas IES se manejan portadas institucionalizadas con el fin de dar cierta homogeneidad a los trabajos realizados por estudiantes y académicos. Se encuentran disponibles en las páginas de internet institucionales.

1.2 Índice o tabla de contenido

Se trata de una tabla donde se desglosan uno a uno los contenidos de las diversas secciones y apartados en las que se divide el reporte. Es importante guardar la coherencia en la organización del índice, ya sea que las diversas partes se clasifiquen por letras, números o tipografías. En algunos casos también se incluyen un índice de tablas y de figuras.

1.3 Resumen, sumario o abstract

Incluye una versión sintetizada de las partes más importantes del reporte, es decir, la fundamentación teórica, el planteamiento del problema, la metodología, los resultados y las conclusiones. Constituye el primer acercamiento que tiene un lector potencial y por tanto se recomienda que sea completo, conciso y claro. La extensión puede variar, pero se recomiendan entre 100 y 200 palabras.

1.4 Introducción

Es la presentación general del trabajo de investigación en la que se definen algunas de sus principales directrices como son las siguientes: antecedentes, planteamiento o definición del problema, objetivos, justificación y limitaciones. En algunos casos también se habla de la importancia o relevancia del estudio para el tema tratado, ya sea porque constituye algo nuevo o porque provee un enfoque diferente al hasta entonces empleado.

En los antecedentes la idea es ubicar al lector partiendo desde un enfoque general hasta uno más particular. El fin es posicionar al objeto de estudio dentro de un referente teórico temporal de fácil comprensión. Esta sección puede abordarse desde dos perspectivas: (a) una más teórica en la que se mencionan estudios realizados previamente; o (b) desde una perspectiva más empírica o temporal, es decir, se enfoca más al desarrollo de un fenómeno existente que probablemente no ha recibido la suficiente atención por parte de los estudiosos. Una combinación de ambos enfoques puede cumplir de manera cabal con el propósito de introducir al lector. Especialmente cuando se trata de un objeto de estudio relativamente nuevo o desconocido.

Por otra parte, en el planteamiento del problema se ubica el foco de estudio. Se describe la situación actual que da surgimiento a la investigación, misma que se puede llevar a cabo desde dos vertientes: por la necesidad de entender un nuevo fenómeno o por la necesidad de una nueva manera de entender un fenómeno previamente estudiado. El punto clave en esta parte lo constituyen las preguntas de investigación general y específicas que guían el proceso y que ya

fueron explicadas anteriormente en el capítulo 1.4 sobre la delimitación de preguntas, el problema de investigación y los objetivos.

En concordancia con las preguntas, se presentan los objetivos, los cuales también se clasifican en general y particulares. Durante la redacción, el tipo de verbos que se eligen para determinar los objetivos son muy importantes, ya que adelantan el tipo de análisis que constituye el informe: descriptivo, causal, experimental, etc. Cabe destacar la necesidad de distinguir entre un objetivo y una estrategia para evitar que se establezca como un fin algo que en realidad más bien define un procedimiento.

Otro apartado relevante de la introducción lo conforma la justificación. Mientras que el planteamiento expone una necesidad, esta parte se refiere más al tipo de contribución que dicho estudio tendrá tanto en la discusión teórica como en el mayor conocimiento de un problema específico. Si se ha problematizado una situación, se requiere mencionar cuáles serán los beneficios que acarreará su estudio con el fin de asegurar una contribución real al conocimiento científico.

Finalmente, en la introducción también se incluyen las limitaciones, es decir el alcance que siempre se debe establecer en todo estudio. En toda investigación hay una serie de propósitos que se expresan en el proyecto. Sin embargo, parte del proceso de crecimiento y maduración de una investigación implica delimitaciones, algunas de las cuales pueden ser completamente inesperadas como recién ocurrió con el surgimiento de la pandemia por Covid-19. Ello obligó a replantear el alcance de muchas investigaciones y es completamente

...en el proceso de investigación

válido y necesario exponer hasta dónde abarca el estudio presentado.

2. El cuerpo del informe

A pesar de las diferencias existentes entre todos los formatos de reportes de investigación, ya sea que se trate de tesis de licenciatura, artículos científicos, ponencias publicadas en memoria y hasta trabajos escolares de breve extensión, hay tres componentes que están siempre presentes. Se trata de la fundamentación teórica, la explicación metodológica, los resultados y su discusión. La distribución de estas partes puede variar ya que, valga la reiteración, no hay un esquema único de organización. En algunos artículos la introducción incluye, aparte de los elementos expuestos anteriormente, la fundamentación teórica. En otros casos, los resultados y su discusión forman dos apartados separados, de los cuales el segundo es a menudo llamado discusión. En otros reportes, la discusión se incluye solo en las conclusiones de manera que pareciera que el apartado de resultados se presentara sin un trasfondo analítico. Prescindiendo de la forma cómo se organiza la información, la triada teoría, metodología y resultados constituye el núcleo de todo reporte de investigación. Lo que se expone a continuación es una propuesta de los elementos constitutivos del cuerpo del informe.

2.1 El marco referencial

Es la parte del reporte que entraña las bases teóricas sobre las que se sustenta la investigación. Implica una

serie de revisiones, reflexiones y propuestas que eventualmente proveen las perspectivas de análisis a partir de las cuales se explica el objeto de estudio. Un marco referencial se compone de dos partes: el marco conceptual y el marco teórico. En el caso del formato solicitado para la experiencia educativa anteriormente mencionada, se incluyó también una revisión de casos similares.

2.1.1 Marco conceptual

Se trata de un recorrido de los principales conceptos en torno al tema de investigación o que de alguna manera contribuyen a entender mejor el fenómeno. En esta parte del marco referencial, es importante hacer alusión constante a la diversidad de ideas presentadas y el objeto de estudio. Esta exploración sirve para agrupar conceptos y hacer comentarios acerca de su validez o alcance para la investigación en curso.

Una parte técnica fundamental de la revisión de literatura la constituye el aparato de citas. Dado que este apartado es de naturaleza expositiva y argumentativa, el uso correcto de un sistema de citación es esencial. Ello requiere que quién investiga esté consciente de las normas que regulan el sistema que decida utilizar. Esto no significa que tenga que memorizar de forma exhaustiva dichas reglas, ya que eso sería imposible, pero sí asegurarse de tener acceso a bibliografía que le permita elaborar de manera adecuada la diversidad de citas que se requieren en la producción escrita académica.

2.1.2 El marco teórico

Uno de los componentes que causa mayor dificultad y confusión al momento de redactar un reporte de investigación es la construcción del marco teórico. A menudo, hay un traslape entre lo que se entiende por marco conceptual. El primero ya fue explicado. Para la construcción del segundo se requieren algunas precisiones. El marco teórico se trata de un conjunto de elementos teóricos (teorías, supuestos o ideas) dentro de los cuales se pueden ubicar los parámetros (dimensiones, categorías) que constituyen un conjunto ordenado. Este sistema tiene como propósito crear un marco de abstracciones estructuradas que permita describir, correlacionar, explicar o interpretar las partes constituyentes de un objeto de estudio. Hay diversas formas en las que este sistema puede ser adoptado. En algunos casos, se trata de teorías completamente pulidas que se aplican a un objeto de estudio nuevo; en otros hay una amalgama de teorías o ideas que se crean porque se considera que son capaces de proveer herramientas para entender el objeto de estudio. En otras circunstancias, un marco teórico puede ser completamente innovador cuando el sistema expuesto parte de conceptos y supuestos desconocidos o poco considerados en un campo disciplinar. En cualquier de estas situaciones, sin embargo, es necesario tener en mente que el marco teórico, como su nombre lo indica, provee los parámetros a partir de los cuales se analizará el objeto de estudio.

Una parte importante de aclarar durante la elaboración del marco teórico es la definición de los conceptos y la relación que guardan entre sí. Este apartado puede

existir de forma separada o puede incluirse dentro del marco conceptual o en el marco teórico. El propósito es que aquellos conceptos que por la relevancia o frecuencia de uso durante el reporte tengan definiciones muy claras con el fin de evitar ambigüedad (por ejemplo, definir el término competencia en un reporte del área de educación). Aquí también es posible encontrar términos de uso común que, dentro del contexto del estudio, adquieren un significado particular que es necesario tener en cuenta para evitar confusiones, como sería el caso del concepto difusión en un texto sobre antropología. De la misma manera, en este apartado también se pueden introducir conceptos nuevos o poco familiares que pueden adquirir nuevos matices de acuerdo con el sentido de la investigación (como cuando se habla de clase invertida en la enseñanza de segundas lenguas). Comúnmente, la definición de estos conceptos implica entender su relación respecto a otros, de ahí que a veces se vuelve necesario aclarar su uso mediante redes semánticas que sirvan de guía al lector para mejor comprensión del reporte.

2.1.3 Revisión de casos similares

Se trata de un ejercicio de exploración de los alcances logrados en investigaciones similares. Por muy novedoso que pueda ser un tema o un enfoque, una investigación no se construye a partir de cero. Hay antecedentes de diversos tipos que pueden estar estrecha o lejanamente relacionados ya sea con la propuesta teórica, metodológica o de objeto de estudio. Al dar a conocer estos autores o estudios, se puede subrayar qué aspectos no han sido atendidos y, por ende, fortalecer la

...en el proceso de investigación

razón de la investigación. También se demuestra que se está consciente de lo logrado hasta ahora para poder ubicar la investigación propia dentro de la configuración disciplinaria. Puesto que se trata de evocar lo realizado, la correcta citación es esencial en este apartado tal como se ha insistido a lo largo de toda la sección sobre el marco conceptual.

3. Metodología

Una vez que ya se presentaron el marco referencial y el planteamiento del problema, es necesario explicar cómo se lleva a cabo la investigación. En esta sección, se da una breve explicación sobre la metodología estadística empleada para lograr los objetivos, es decir se describe el proceso que tiene lugar en la obtención de resultados que dan respuesta a las preguntas de investigación. La narrativa utilizada es en tiempo pasado, pues describe los logros, desafíos y experiencias encontradas.

3.1 Diseño estadístico

La primera parte de la metodología con énfasis cuantitativo se refiere al diseño estadístico. El diseño estadístico se puede entender como “la guía que conduce todo el proceso” (Ojeda Ramírez et al., 2011, p. 39), como una hoja de ruta. Por este motivo en este apartado se habla no sólo de lo que se hizo sino también cómo se hizo. Se inicia con la descripción del problema de investigación y por qué fue necesario utilizar una metodología con énfasis cuantitativo. Para ello, se debe explicar quiénes constituyen las unidades de estudio, el proceso mediante el cual se seleccionaron, así como el

tipo de variables que se decidió considerar. Otro aspecto que hay que destacar en la redacción es el tipo de estudio que se llevó a cabo. Definir si es de tipo descriptivo o analítico, y si se realizó con una muestra o se hizo seguimiento temporal, es decir si es longitudinal o transversal. También es necesario dejar claro el tipo de muestreo que se usó y qué características y tamaño tenía la muestra junto con los diversos tipos de ajustes que se hayan presentado. La redacción de esta parte del proyecto se realiza utilizando los verbos del tiempo pasado.

3.2 Análisis estadístico

El apartado de análisis estadístico tiene como propósito describir el proceso al que se someten los datos. El fin es obtener resultados que respondan a las preguntas de investigación. Tras una breve introducción, se inicia con la descripción de la herramienta utilizada, lo cual es generalmente un programa estadístico como SPSS, R u Orange y se pueden mencionar los criterios de su elección. Posteriormente, se da paso a las diferentes fases del análisis y cómo se llevan a cabo. Por ejemplo, en la fase descriptiva o exploratoria hay que destacar los aspectos generales extraídos del análisis marginal, que se recomienda sea exhaustivo. Ello permitirá un análisis bivariado o cruzado más enriquecido (Ojeda Ramírez et al., 2011). Si se utilizan técnicas multivariadas también es necesario profundizar en su utilidad para los fines de investigación. Como en el apartado anterior, la redacción se hace con verbos en el pasado y se deben explicar no solo los diversos tipos de análisis, sino también cómo se

...en el proceso de investigación

hicieron y el fin de cada de ellos. Para el lector es prioritario entender el objetivo de los diversos tipos de análisis y cómo se relacionan con los objetivos establecidos en el estudio.

4. Resultados y discusión

Si comparáramos el reporte con una comida de tres tiempos, diríamos que hemos llegado al plato fuerte. Esto no significa que lo anterior no sea importante, sino que ya una vez que se ha preparado el terreno mediante las bases teóricas y metodológicas, se da paso a aquello que dará respuesta a las preguntas de investigación y establecerá si los objetivos han sido alcanzados. Como su nombre lo indica, esta sección consta de dos apartados: resultados y discusión. La parte de resultados se refiere a la presentación ordenada que se hace de los datos posteriores al análisis estadístico. Algo que los especialistas destacan es evitar la confusión de resultados con discusión, donde ya hay un elemento analítico y de interpretación, como se ve más adelante.

4.1 Resultados

El apartado de resultados generalmente consiste en dos partes: el análisis descriptivo y el análisis inferencial. Puede iniciar con un panorama general de los resultados más destacados para posteriormente profundizar. En el formato que se maneja durante el Seminario, se pide de entrada en el apartado de resultados el desglose de objetivos general y particulares. Ello sirve de guía para revisar con detalle el cumplimiento de cada uno de los objetivos. Posterior a ello, se presentan los resultados

emanados del análisis exploratorio, primeramente. Con esto, es posible tener un panorama general de cada una de las variables consideradas en el estudio. Por cada variable se muestran tablas y figuras que representan de manera visual los resultados. Posteriormente, de forma narrativa se hace una explicación más detallada con el fin de tener amplia comprensión de los datos que ahí se presentan.

La segunda parte del apartado la constituye el análisis bivalente y multivalente. Al igual que los resultados mostrados previamente, la información aquí también se organiza por objetivos. De esta manera, los análisis bivariantes consisten en el cruce entre cada variable dependiente y la variable independiente. Estos objetivos también van acompañados de gráficos de apoyo como pueden ser tablas de frecuencia y gráficos de barras. En el caso del análisis multivariado se presentan los resultados extraídos de la técnica multivalente utilizada (correlación canónica, componentes principales, análisis de agrupación o de correspondencia, entre otros). Tal como se hace en la parte anterior, los gráficos van acompañados de textos narrativos donde se destacan los principales resultados y se mencionan aquellos detalles que puedan llamar la atención por sus características particulares.

Antes de proceder al apartado de discusión vale la pena detenerse un poco para hacer algunos comentarios sobre el análisis estadístico y el uso de tablas y gráficos. Sobre el primero, se dice que los análisis estadísticos pueden ser tan variados como la imaginación lo permita. Giroux y Tremblay (2004) mencionan diversas opciones: resultados que se pueden descomponer tanto como sea posible, búsqueda de relaciones entre variables para

...en el proceso de investigación

obtener resultados más matizados, establecimiento de contrastes y estudios de los casos que no se ajustan a la norma, por ejemplo. En lo concerniente a las tablas y gráficos, algunos especialistas (Hernández Sampieri et al., 2014; Ojeda Ramírez et al., 2011) han emitido algunas recomendaciones acerca de la presentación de las tablas y gráficos. Si bien el reporte de evidencias puede ser exhaustivo, ya que se debe explorar cada uno de los objetivos planteados, la recomendación es que el reporte solo contenga aquellas tablas o gráficos que sean de mayor relevancia y que respondan directamente a las preguntas de investigación. En cuanto a las tablas estadísticas tratadas de forma individual, las recomendaciones tienen que ver con la claridad de su presentación: título conciso y relacionado con el contenido, datos estadísticos pertinentes y en cantidad manejable, organización lógica, leyendas claras y estricto apego a las normas del manual de citación por el que el autor se haya decantado o se le haya solicitado.

4.2 Discusión

La segunda parte de esta sección es la discusión, es decir el momento de confluencia de los diversos componentes del proceso de investigación. A diferencia de su predecesor, la discusión va más allá de la mera presentación de resultados pues requiere la necesaria confrontación que tiene que haber entre los resultados emanados del análisis estadístico y el marco teórico con el fin de evaluar si las respuestas de investigación han sido respondidas. Este movimiento tiene un doble sentido: si bien es verdad que se ponen a prueba los resultados, también es cierto que se prueba la solidez de

la teoría propuesta para explicar el fenómeno. Ello puede derivar en una propuesta de ajustes necesarios o, si los resultados difícilmente se ajustan al marco elegido, aventurar una explicación completamente nueva. Para lograr este fin, en este apartado es válido hacer un contraste con los resultados obtenidos en investigaciones anteriores.

Es posible que la discusión se refiera a algo más que sólo los resultados. Giroux y Tremblay (2004) sugieren una serie de temas que pueden ser origen de problemas durante el desarrollo de la investigación y que, por tanto, se pueden incluir en este apartado: la operativización de los conceptos, la selección de las técnicas de recolección de datos, el grado de pertinencia del muestreo, la claridad de las preguntas del instrumento, el instrumento mismo y el desarrollo de la investigación. Todos estos elementos ayudan a entender mejor los resultados, a cuestionar la validez de algunos, a matizar posiciones e incluso, declararlos no aptos para considerar que un objetivo ha sido alcanzado.

La mención de aspectos problemáticos no significa que se deba restar validez a la investigación. A menos que se trate de un cuestionamiento al modelo en absoluto por aspectos de etapas no desarrolladas propiamente, por ejemplo, en realidad es imposible hablar de investigaciones sin defectos. Siempre hay aspectos que mejorar, aspectos en los que quizá se debió abundar, técnicas que se pudieron usar en lugar de otras, etc.

4.3 Conclusiones

En redacción académica ninguna sección crea mayor disenso que las conclusiones. Mientras que para algunos se trata sólo de una recuperación somera de los principales resultados, para otros la discusión debería estar presente aquí y no en la sección anterior, además de incluir otros rasgos. En voz de Sabino (1998), en las conclusiones hay “un cierto elemento de libertad”. La presente propuesta se enfoca tanto a elementos del estudio como a otros no académicos. Tradicionalmente, en las conclusiones se resaltan los aspectos más relevantes del estudio desde una perspectiva global. También es una parte en la que pueden venir a colación limitaciones que futuras investigaciones deben considerar, o hacer propuestas de hilos sueltos que salieron a la luz y en las que por cuestiones de tiempo y espacio no se pudo profundizar o quedaron fuera del objetivo.

Las conclusiones pueden abarcar aspectos que van más allá de lo meramente académico. Una vez que los resultados han sido dados a conocer y discutidos, quien investiga se puede aventurar a propuestas relacionadas con los sujetos involucrados en su estudio: prácticas, iniciativas, políticas inclusive, que se dirijan directamente a corregir aquellos problemas descubiertos durante el proceso. No obstante, es preciso hacer una cuidadosa anotación en esta última parte. Un reporte de investigación puede plantearse el objetivo de comprender un problema, descifrar cómo se configura o tratar de identificar sus causas. El objetivo nunca es ofrecer la solución. No sólo se trata aquí de una cuestión de modestia intelectual, sino que la toma de decisiones corresponde a otra esfera que, aunque puede apoyarse de éste, es ajena al quehacer intelectual.

5. Referencias

La elaboración del listado final de referencias va de la mano con las obras citadas a lo largo de todo el reporte. Como se explicó en el apartado Marco conceptual, el correcto uso del sistema de citado es esencial, no sólo durante la redacción del cuerpo principal sino también en el listado de referencias. Ambos elementos son indisolubles uno del otro: todas y cada una de las obras que aparecen citadas en el cuerpo del texto deben encontrarse en el listado de referencias y, a la inversa, el listado de referencias no puede incluir más que las obras que han sido citadas en el cuerpo de referencia. Esto no sólo facilita a los lectores la búsqueda de las referencias completas al momento de leer el texto, sino que también demuestra solidez y seriedad del trabajo.

Es importante no confundir las referencias con la bibliografía. Mientras que la primera contiene sólo las obras citadas, se considera que en la segunda se pueden incluir aquellas obras que se consultaron y pudieron servir de sustento para el trabajo sin que se hayan incluido alguna cita en específico. La bibliografía también puede servir de orientación al lector en caso de que se quiera abundar sobre otros temas que no se profundizaron en el reporte.

Un listado de referencia exitoso implica necesariamente el uso de un manual de estilo. Si bien existen diversos estilos (Chicago, Harvard, MLA, Vancouver), en el área de las Humanidades y Ciencias Sociales goza de amplio uso el publicado por la Asociación Americana de Psicología, mejor conocido como Manual APA (por sus siglas en inglés). En el caso

...en el proceso de investigación

de los trabajos que se solicitaron como parte del Seminario, el manual de estilo utilizado corresponde a este último, en su edición séptima; es decir, se usó APA 7, aunque en este libro se utiliza el estilo Harvard. Existen varios recursos de apoyo para el uso correcto del Manual APA. La obra original se publica originalmente en inglés y, por este motivo, la mayoría de los recursos disponibles en línea se encuentran en este idioma. La página oficial es <https://apastyle.apa.org>. En español se puede encontrar una versión resumida con los aspectos más importantes sobre las referencias en la siguiente página: <https://normas-apa.org>. Otro recurso de apoyo puede ser la obra de De Volder et al. (2020), la cual contiene algunas de las normas básicas para citación y elaboración del listado de referencias.

6 . Apéndices

Por último, se incluyen los apéndices. Se trata de documentos que sirven de apoyo adicional al reporte y se colocan en la parte final después de las referencias. A lo largo del texto es posible que se haga mención a elementos que estén directa o indirectamente relacionados con la investigación, pero que son hasta cierto punto adicionales al tema principal. No obstante, con el fin de apoyar al lector interesado se proporciona más información o mayor detalle sobre el tema, sin que ello suponga una interrupción de la lectura principal.

Los apéndices se colocan conforme su orden de mención en el texto y se les asigna una letra o número (Apéndice A, Apéndice B; Apéndice 1, Apéndice 2, etc.). Los tipos de documentos que comprenden los apéndices

son muy variados. A continuación, se mencionan algunos ejemplos:

- Documentos que por su relevancia ameritan ser incluidos de forma completa, ya sea por su contenido histórico, legal, científico, etc. o porque se utilizaron en la investigación, como puede ser el cuestionario utilizado en la encuesta.
- Documentos de tipo técnico como puede ser una descripción más detallada del programa de paquetería utilizada para el cálculo estadístico o el tipo de equipo usado.
- Datos de tipo estadístico o matemático que requieran una lectura más detallada y conocedora del alcance que tenga el reporte. En ocasiones, los cálculos que se hayan realizado para el muestreo o la validación de instrumentos también se incluyen aquí.
- Diversos tipos de material gráfico (dibujos, fotografías) que puede ser ilustrativo, pero cuya relevancia no es central en el reporte.

A manera de conclusión

El proceso de investigación inicia con la concepción de una idea y culmina con la redacción del reporte. Como se explicó al inicio, esta última etapa no es lineal y se compone de avances y retrocesos. Por tal motivo, la planeación adaptativa es importante. La corrección es otra parte esencial de la redacción, lo cual en ocasiones puede tornarse cansado. Si ese es el caso, ello amerita dejar que el texto repose un momento, no demasiado, sólo lo suficiente para aclarar los pensamientos y retomar las ideas. Una vez que se tenga lograda la primera versión, vale la pena alejarse un par de días

...en el proceso de investigación

para hacer una lectura completa del texto con una nueva mirada. Esto permitirá que veamos con nuevos ojos errores de escritura, repeticiones, redundancias, ambigüedades, etc., cuya corrección se reflejará en un mejor texto. Otra recomendación es que una vez lograda la versión final se comparta con personas allegadas con el fin de conocer otros puntos de vista que vean más allá del nivel de auto saturación que en ocasiones alcanza quien redacta. Tras incorporar esas observaciones, podemos decir que el reporte está listo para ser presentado.

Referencias

- Amelia Acosta León, *Del proyecto a la tesis. Guía metodológica para la realización de trabajos de investigación*, México, Universidad Juárez Autónoma de Tabasco, 2012.
- Fidias Arias, *El proyecto de investigación. Introducción a la metodología científica*, Venezuela, Episteme, 2012.
- Carolina de Volder, Ignacio Mancini y Lucia Marpegan, *La cita documental. Elementos y ejemplos de referencias en estilo APA*, Argentina, Instituto de Investigaciones Gino Germani, 2020. Consultado en: https://iigg.sociales.uba.ar/wp-content/uploads/sites/22/2020/11/LaCita4taed_10_11_20.pdf
- Sylvain Giroux y Ginette Tremblay, *Metodología de las ciencias humanas, La investigación en acción*, México, Fondo de Cultura Económica, 2004.

- María del Carmén Gómez Chagoya, *Unidad 4. Reporte de investigación*, México, Universidad Autónoma del Estado de México, 2004.
- Roberto Sampieri, Carlos Collado y María del Pilar Baptista, *Metodología de la investigación*, México, Mc.Graw Hill, 2014.
- Mario Miguel Ojeda Ramírez, José Enrique Díaz Camacho, Clara Apodaca Victoria y Isarel Trujillo Landa, *Metodología del diseño estadístico*. México, Universidad Veracruzana, 2011.
- Carlos Sabino, *Cómo hacer una tesis y elaborar todo tipo de escritos*, Argentina, Lumen-Humanitas, 1998.
- Edgar Antonio Tena Suck y Bernardo Turnbull Plaza, *Manual de investigación experimental. Elaboración de tesis*, México, UIA-Plaza y Valdés México, 2007.

Cultura estadística en la investigación

EXCIANI ADUY ALARCÓN SANTAMARÍA
MARIO MIGUEL OJEDA RAMÍREZ

Introducción

Antes de iniciar se debe considerar que la estadística es una parte importante de las disciplinas que nos ayudan a comprender y aprender del mundo en el que vivimos. Nos proporciona herramientas para estudiar fenómenos de los que podemos obtener datos o evidencias. Provee de métodos y técnicas de análisis y proporciona principios para interpretar, inferir y elaborar juicios o conclusiones sustentadas. En términos más simples se puede decir que la estadística es la disciplina científica que nos dota de métodos y técnicas mediante los cuales se obtienen los datos, proporciona los métodos de análisis y guía en la presentación de los resultados. Sin olvidar que estos deben responder al objetivo de una investigación o estudio fáctico; es decir, de un estudio que está basado en hechos (Ojeda, 1994).

En la sociedad de la información y el conocimiento, permeada por las TIC, la estadística está tomando gran relevancia. Se ha hecho necesario procesar datos para obtener conocimiento y darlo a conocer a la sociedad; de esta manera es posible tomar decisiones fundamentadas. Los ciudadanos reciben información de

distintas maneras y de diferentes fuentes, por lo que requieren ser competentes para interpretar esas informaciones. Desde el nivel educativo básico, los conceptos de la estadística se han integrado como una parte importante de los contenidos, para lograr cimentar en los estudiantes una cultura estadística (Gal, 2002). En este capítulo se tratará sobre la cultura y el pensamiento estadísticos en el proceso de investigación.

Definición de la cultura estadística

¿Qué es la cultura estadística? Wallman (1993) la define como

la habilidad de comprender y evaluar críticamente los resultados estadísticos que permean nuestra vida diaria, complementándose con la habilidad de apreciar las contribuciones que el pensamiento estadístico puede hacer tanto a nivel público como privado, en la toma de decisiones profesionales y personales. (p. 1)

En esta cita es posible observar cómo la cultura estadística es relevante para todos los individuos que conforman a la sociedad. A partir de ella pueden consumir críticamente la información que las diferentes organizaciones, gubernamentales y no gubernamentales, dan a conocer sobre los acontecimientos políticos, económicos, sociales, educativos y de salud que acontecen en distintos sectores de la sociedad. Asimismo, se podrá juzgar y entender cómo con la estadística estas instituciones sustentan la toma de decisiones para hacer frente a los diferentes asuntos y problemas que afectan a la población.

...en el proceso de investigación

Por ejemplo, en la pandemia de Covid-19 han sido los resultados estadísticos los que refieren el estado actualizado de la población contagiada, los hospitalizados, los decesos, etc. En el caso de la Organización Mundial de la Salud (OMS), su sistema mundial de estadísticas generó datos y representaciones de los mismos (gráficos), al cual tienen acceso todas las personas interesadas en el tema; mediante estos datos se explican la variación, la distribución y las correlaciones que se han generado en diversos países y al interior de éstos, brindando reportes para que los gobiernos y los ciudadanos actuaran en consecuencia. Sin embargo, para explorar y comprender estos gráficos es necesario que la población posea una cultura estadística o *statistical literacy*. Gal (2002) la menciona como una habilidad clave en la sociedad de la información; considerada parte de la herencia cultural necesaria en cada ciudadano. Este autor señala que es necesario desarrollar en el individuo dos capacidades: la de interpretar y evaluar críticamente, y la de comunicar opiniones sobre la información que proviene de la estadística. Además, distingue entre conocimientos básicos como la alfabetización estadística, y un conocimiento más profundo que implica un razonamiento abstracto, el contexto y el desarrollo del pensamiento crítico. Todo lo cual es necesario para lograr una correcta cultura estadística en la sociedad de la información y el conocimiento.

La transnumeración

El desarrollo de las capacidades de lectura de informaciones emanadas de datos está mediado por un

proceso llamado transnumeración, término acuñado por Wild y Pfannkuch (1999). Se entiende como la transformación numérica para facilitar su comprensión en relación con los fenómenos del mundo real que describen. Este proceso es de vital importancia en la cultura estadística. Permite la lectura y la interpretación de la información derivada de los resultados de procesos estadísticos, para comprender una situación en el contexto real. Esta transnumeración propicia un pensamiento denominado transnumerativo; la posesión de este pensamiento queda evidenciado en el momento en el que el ciudadano o profesional puede leer e interpretar una representación estadística correctamente. Un ejemplo, sería entender el ordenamiento o el agrupamiento de entidades descritas por indicadores, leer frecuencias, proporciones, medidas de tendencia central, medidas de dispersión, de correlación, de asociación, entre otras. O bien dar tratamiento a los datos y las gráficas o tabulaciones para comunicar resultados.

Chick et al. (2005) establecen tres fases para el desarrollo del pensamiento transnumerativo: (a) captura de datos del mundo real considerando el contexto y los instrumentos; (b) reorganización y cálculos con datos a partir de lo que se quiere comunicar; y (c) la comunicación del mensaje mediante una representación estadística. En educación básica la transnumeración se genera en un ambiente didáctico con actividades cotidianas. Por ejemplo, cuando los niños recogen información (datos) con los estudiantes de la escuela indagando ¿cuál es el sabor de su helado favorito? Después agrupan y clasifican los datos. Más tarde generan gráficas que representen lo que han

encontrado; por último, comunican sus hallazgos. Con ello se busca que conozcan, comprendan y produzcan información estadística para conocer y entender su realidad (González-Ruiz et al., 2015).

En la educación superior la transnumeración se expresa en la investigación fáctica de énfasis cuantitativo, principalmente. El método científico tiene una estructura metodológica en la que se requiere de habilidades estadísticas para identificar qué se va a investigar; es decir, la elección de un tema que sea interesante para una audiencia y que aporte conocimiento. Lo anterior requiere planear un buen diseño de la investigación. Para evitar errores en el proceso de recolección y análisis se deben (a) tener objetivos claros; (b) para qué servirá investigar el tema elegido; y (c) establecer de manera idónea cómo se medirá. Es decir, que instrumentos de recolección de datos se utilizarán teniendo presente las variables necesarias para lograr análisis de todo tipo, incluyendo los multivariados que permitan obtener conclusiones (Ojeda, 1994).

De acuerdo con la Organización de las Naciones Unidas (ONU, 2012) la cultura estadística describe la capacidad de los ciudadanos de manera individual o grupal para entender la estadística. También señala que la cultura estadística requiere de habilidades matemáticas y conocimientos básicos de estadística. Gal (2002) indica que para poseer un grado de cultura estadística convergen varios factores como la formación académica, la alfabetización, el conocimiento del mundo y las creencias individuales con respecto a las matemáticas básicas y la estadística.

Características de la cultura estadística

Actualmente los estadísticos consideran que la comprensión de la metodología estadística es primordial en varios ámbitos de la sociedad. Por ello aumentar la cultura estadística en el público general se está volviendo una prioridad para enseñarles a comprender el mundo que les rodea (ONU, 2012). La ONU distingue tres fases para desarrollar la cultura estadística:

1. La alfabetización estadística. Tiene que ver con la necesidad de contar con un conocimiento estadístico para comprender datos cuantitativos y hechos. Quiénes no poseen este conocimiento entran en una categoría de analfabetismo estadístico. Para evitarlo se requiere manejar las matemáticas con el objetivo de reconocer porcentajes, promedios y estimaciones.
2. La transmisión estadística. Consiste en leer y comunicar los resultados de los datos de manera clara, con calidad, con explicaciones contextuales que ayuden a los usuarios a comprender las representaciones gráficas.
3. Descubrir el uso de la estadística en la toma de decisiones personales y profesionales. Esto mediante una correcta comprensión de las cifras expresadas en la información a la que acceden los usuarios.

Así mismo, la ONU realiza una clasificación de tres niveles de cultura estadística: (a) nivel básico, son personas que tienen la capacidad de identificar términos estadísticos y leer representaciones gráficas; (b) nivel medio, identifica los métodos que utilizan las investigaciones o comunicaciones, entiende definiciones y medidas estadísticas; y (c) nivel avanzado, comprende la terminología estadística, detecta errores de muestreo,

conoce los elementos que conforman un estudio estadístico, hace inferencias y evalúa críticamente los datos. En este último nivel se encuentran personas versadas en los fundamentos de la estadística, en su filosofía y en su historia.

Los usuarios de la cultura estadística pueden ser (a) estudiantes, para que entiendan y utilicen de manera eficaz la información que proporcionan los datos; (b) profesores que fomenten la enseñanza de la estadística dentro del currículum; (c) instituciones de enseñanza superior, cuyo fin es que sus egresados tengan un perfil que les permita entender, analizar, interpretar y evaluar información estadística; (d) tomadores de decisiones en instituciones gubernamentales y no gubernamentales, para que logren objetivos enfocados a un bien común; y (e) líderes de opinión, para que comuniquen de manera correcta y eficiente la información estadística que permita comprender los hechos que acontecen en el mundo.

La enseñanza de la estadística es una de las acciones relevantes que se deben considerar para desarrollar la cultura estadística. Roiter y Petrocz (1996) mencionan cuatro paradigmas: pensar a la estadística no como una rama de las matemáticas, como análisis de datos, como planificación estratégica y como una asignatura que se aprende planteando y resolviendo problemas de índole estadística.

El razonamiento abstracto y el dominio de las matemáticas básicas es una base importante para aprender, comprender y realizar estadística. El segundo se refiere a la función principal de la estadística. Es decir, realizar análisis de datos (éstos se presentan en la vida cotidiana) para comprender una situación mediante representaciones gráficas que den a conocer información

interesante. El tercero se enfoca en planear una estrategia para abordar un estudio. Esto es fundamental, pues guía cualquier investigación y atiende a tiempo cualquier eventualidad. Por último, se encuentra el aprendizaje basado en problemas (ABP), el cual propicia un ambiente idóneo para emplear la estadística vinculada con la vida real (Ojeda, 1994; Cano-Orellana, 2022).

En este tenor Goded (2015) resalta que el conocimiento estadístico y probabilístico siempre debe partir de las problemáticas que rodean al estudiante. Esto permitirá generar en ellos una competencia social que coadyuve en la construcción de su propio conocimiento a través del saber hacer. Todo esto propicia un aprendizaje significativo (Ausubel, 1978), relacionando la nueva información con conocimientos previos. A su vez logra incorporar los nuevos conocimientos a su estructura cognitiva.

Contreras y Molina-Portillo (2019) describen seis componentes de la cultura estadística. El primero se refiere a la comprensión, interpretación y argumentación de la información obtenida por metodología estadística. Hace referencia al entendimiento de la terminología estadística y a las habilidades lingüísticas para reflexionar, interpretar y comunicar esta información. El segundo abarca la actitud crítica y el cuestionamiento. Este componente está permeado por las creencias y emociones del individuo como ser social, y su habilidad para descubrir a través de la evaluación crítica las estructuras ocultas en la información. El tercer componente es la detección de sesgos y errores. Tiene que ver con la capacidad de los consumidores o individuos de detectar errores en la información que

brindan los medios de comunicación. El cuarto es el contexto. Es un elemento importante puesto que permite una claridad del lugar, modo, tiempo y espacio en el que se producen los datos para interpretarlos y comprenderlos. El quinto contempla las destrezas en matemáticas básicas. Va en relación con los conocimientos y habilidades numéricas. Por último, la transnumeración, mediante la utilización de representaciones gráficas que permitan la comprensión de la información. En conjunto, estos componentes permiten en el individuo una cultura estadística sólida.

La cultura estadística dentro de la investigación está relacionada con conocer los principios, objetivos y algunas técnicas de carácter descriptivo. Esto le permite a quien investiga un manejo básico de la información numérica e identificar los problemas que requieren de un especialista en estadística. Las investigadoras y los investigadores requieren de la cultura estadística en todos los estudios. A este respecto Ojeda (1994) enumera cinco momentos: (a) en la revisión de casos similares para elaborar un estado del arte o de la cuestión; (b) en la delimitación de objetivos y plan de trabajo; (c) en la obtención de los datos; (d) en el análisis e interpretación de los resultados; y (e) en la elaboración del reporte de investigación. Como se puede observar la cultura estadística está de inicio a fin en una investigación, de ahí su importancia. Este autor señala que la cultura estadística genera las siguientes habilidades que son de gran importancia en su quehacer académico:

1. Entender que la obtención de datos es una acción de medición; por lo tanto, debe definir qué estudiará,

- cómo lo realizará y qué instrumento utilizará para medir aquello que pretende estudiar.
2. Comprender que los datos le aportan información para describir colectivos. Esto es que partirá de la unidad para inferir y posiblemente generalizar.
 3. Identificar en los colectivos ciertos patrones que le permitan estudiar comportamientos atípicos.
 4. Comprender que las inferencias que pueda lograr a partir de los datos siempre estarán sujetos a incertidumbre.
 5. Interpretar resultados, diseñar representaciones gráficas para comunicar los hallazgos y producir conclusiones.

Con estas habilidades quienes investigan contarán de los elementos necesarios para diseñar, realizar y comunicar los resultados de nuevas investigaciones en diversos temas, ya que la investigación comprende a todas las áreas académicas. El método científico cuenta con una serie de etapas.

1. Motivación y antecedentes. Es fundamental que quién realiza la investigación tenga un interés especial en el tema a investigar. La investigación le llevará un tiempo y varias actividades a desarrollar. Posterior a ello, es importante investigar los antecedentes de la temática, buscar que se ha realizado al respecto, cómo han abordado otros autores el mismo tema para que la contribución que se realice sea relevante.
2. Objetivos, tanto general como específicos de la investigación. Es decir, tener claro qué se pretende lograr y con ello guiarse durante el proceso.
3. Hipótesis y variables.
4. Variables operativas, procedimentales y escalas de medición. Si se tiene claro este dato será más fácil

...en el proceso de investigación

- elegir y elaborar los instrumentos de recolección de datos.
5. Procedimiento y fuentes de datos. Éstas últimas pueden ser primarias (mediante encuestas o instrumentos aplicados por quién investiga) o secundarias (archivos, censos, bases de datos generados por alguien más).
 6. Análisis. Se realiza con el apoyo de un software adecuado a los datos que se obtienen y a los objetivos que se pretenden alcanzar. La o el investigador analizará los datos realizando varios tratamientos para producir información que sea susceptible para generar conclusiones.
 7. Informe de resultados. Este debe tener una lógica, debe ser claro, sucinto y estar organizado de tal manera que exponga todas las partes de la investigación.

Por tanto, la cultura estadística es de gran importancia para investigar en cualquier disciplina. Permite identificar las etapas y tareas que son de su competencia dentro de la investigación y cuáles ya no entran en su labor. Con ello expresa un grado de cultura estadística que le permite reconocer cuando dar paso a los profesionales en estadística. Son los especialistas quienes deben (a) encargarse del diseño de la base de datos; (b) la verificación de la calidad de datos en la captura, buscando minimizar errores que generen controversias al momento del análisis; (c) la selección del paquete estadístico apropiado, considerando que existe una variedad de software especializado cuyo uso depende del tipo de análisis. Esto permitirá una mayor contribución al conocimiento; (d) la ejecución de las exploraciones y análisis preliminares sacando

frecuencias; (e) el diseño de una estrategia de análisis; y (f) la implementación de un análisis estadístico acorde.

Conclusión

Podemos concluir, en primer lugar, que la cultura estadística es uno de los conocimientos generales que cualquier persona debería poseer para interpretar los acontecimientos que influyen en su vida cotidiana. En segundo lugar, la enseñanza de la metodología estadística es de vital importancia en la educación. Debe ser incorporada al currículum de todos los niveles educativos mediante la transnumeración, para desarrollar habilidades estadísticas que ayuden al desenvolvimiento adecuado en un mundo de números y gráficas. En tercer lugar, la enseñanza de la estadística debe ser uno de los pilares de la educación. Debe estar alineada al saber hacer, lo que se puede lograr mediante el aprendizaje basado en problemas al situar a los estudiantes en la vida real.

Referencias

David Paul Ausubel, *Psicología educativa. Un punto de vista cognitivo*, México, Trillas, 1978.

Antonio Cano-Orellana, “Avanzando hacia una cultura estadística para economistas. Ciclo de Mejora en el Aula de Estadística Avanzada (cómo hacer comprensible y útil la estadística aplicada a la Economía)”, en *Ciclos de mejora en el aula. Año 2021. Experiencias de innovación docente de la Universidad de Sevilla*, Rafael Porla, Elisa Navarro

...en el proceso de investigación

- y Ángel Villarejo (coord.), España, Universidad de Sevilla, 2022, pp. 527-548.
- Helen Chick, Maxine Pfannkuch y Jane Watson, “Trans-numerative thinking: finding and telling stories within data”, *Curriculum Matters*, núm. 1, 2005, pp. 87-108,
- José Miguel Contreras y Elena Molina-Portillo, “Elementos clave de la cultura estadística en el análisis de la información basada en datos”, en *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*, José Miguel Contreras., María Magdalena Gea, María López-Martín y Elena Molina-Portillo (eds.), España, Universidad de Granada, 2019, pp. 1-12.
<https://digibug.ugr.es/handle/10481/55035>
- Iddo Gal, “Adults’ statistical literacy: Meanings, components, responsibilities”, *International Statistical Review*, núm. 1, 2002, pp. 1-51.
- Pilar Goded, “Los escenarios de aprendizaje. Una estrategia para tratar los conocimientos estocásticos en las aulas”, en *Didáctica de la estadística, probabilidad y combinatoria*, José Miguel Contreras, Carmen Batanero, Juan Godino, Gustavo Cañadas, Pedro Arteaga, Elena Molina, María Magdalena Gea y María del Mar López (ed.), España, SEIEM, 2015, pp. 69-86.
- Ignacio González-Ruiz, Teresa González y Myriam Codes, “Análisis de la construcción de la definición de la estadística por maestros en formación inicial”, en *Didáctica de la estadística, probabilidad y combinatoria*, José Miguel Contreras, Carmen Batanero, Juan Godino, Gustavo Cañadas, Pedro Arteaga, Elena Molina, María Magdalena Gea y

- María del Mar López (ed.), España, SEIEM, 2015 (pp. 109-115).
- Mario Miguel Ojeda Ramírez, “La importancia de una buena cultura estadística en la investigación”, *La Ciencia y el Hombre*, núm. 17, 1994, pp. 143-152.
- Cómo hacer comprensibles los datos. (Organización de las Naciones Unidas, 2012). Consultado en: https://unece.org/DAM/stats/documents/writing/MDM4_Spanish.pdf
- Katrina Roiter, & Peter Petrocz, “Introductory statistics courses—A new way of thinking”, *Journal of Statistics Education*, julio de 1996 ,núm. 2. Consultado en: <http://jse.amstat.org/v4n2/roiter.html>
- Katherine Wallman, “Enhancing statistical literacy: Enriching our society”, *Journal of the American Statistical Association*, núm. 421, 1993, pp. 1-8.
- Christine Wild, & Maxine Pfannkuch, “Statistical thinking in empirical enquiry”, *International Statistical Review*, núm. 3, 1999, pp. 223-265.