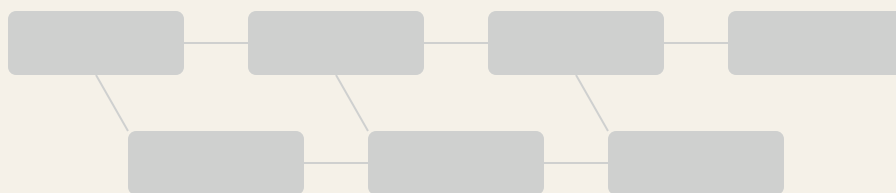


Metodología Experimental en Psicología

Guía para Estudiantes de Grado



Metodología Experimental · Grado en Psicología

Nicolás Sánchez-Álvarez

Departamento de Psicobiología y Metodología de las Ciencias
del Comportamiento · Universidad de Málaga

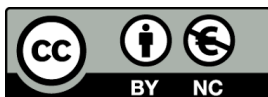
© UMA Editorial. Universidad de Málaga
Bulevar Louis Pasteur, 30 (Campus de Teatinos) - 29071
Málaga www.umaeditorial.uma.es

© Los autores

Diseño y maquetación: Los autores

ISBN: 978-84-1335-487-3

Publicado en mayo de 2026



Esta obra está sujeta a una licencia Creative Commons:
Reconocimiento - No comercial - (cc-by-nc):
<https://creativecommons.org/licenses/by-nc/4.0/deed.es>
Esta licencia permite a los reutilizadores distribuir,
remezclar, adaptar y desarrollar el material en cualquier
medio o formato únicamente con fines no comerciales y
siempre que se otorgue la atribución al creador.

Metodología Experimental en Psicología:

Guía para Estudiantes

Contenido

Contextualización científica de la Metodología Experimental.....	4
1. El método científico.....	4
2. La investigación en la Psicología.....	10
3. Metodología de Investigación en Psicología.....	14
3.1. Metodologías de análisis bibliográfico y/o documental.....	17
3.2. Metodologías empíricas cuantitativas.....	19
3.3. Metodologías empíricas cualitativas.....	24
3.4. Metodologías mixtas.....	28
3.5. Metodologías de Síntesis.....	29
4. El control de las variables extrañas.....	33
5. Diseños de investigación.....	39
5.1. Diseños con estrategia manipulativa.....	41
5.1.1. Diseños experimentales.....	41
5.1.2. Diseños cuasi-experimentales.....	42
5.2. Diseños con estrategia no manipulativa.....	46
6. Validez.....	49
6.1. Validez de Constructo.....	51
6.2. Validez Interna.....	52
6.3. Validez Externa.....	57
6.4. Validez de Conclusión Estadística.....	59
7. Modelos de Análisis Estadístico.....	62
7.1. Modelo Lineal Clásico.....	66
7.2. Modelo Lineal Generalizado.....	68
7.3. Modelo Lineal Mixto (Nivel avanzado).....	69
7.4. Modelo Lineal Mixto Generalizado (Nivel avanzado).....	70
7.5. Modelo de Ecuaciones Estructurales (Nivel avanzado).....	71
Conclusiones.....	73
Referencias.....	74

Contextualización científica de la Metodología Experimental

1. El método científico

Toda ciencia parte de preguntas. En Psicología, esas preguntas giran en torno al comportamiento humano: ¿Por qué recordamos unas cosas y olvidamos otras? ¿Qué hace que una persona sea más resiliente que otra? ¿Cómo influye el entorno en nuestras decisiones? Para responder a estas preguntas de forma rigurosa, es decir, con evidencias verificables y no solo con opiniones o intuiciones, la Psicología utiliza el método científico. Este apartado explica en qué consiste ese método y por qué es la herramienta fundamental del psicólogo investigador. La ciencia es un cuerpo de conocimiento sistemático y verificable que investiga, clarifica y pronostica los fenómenos sociales, artificiales y naturales. El conocimiento científico se adquiere de manera metodológica a través del proceso de observación y experimentación en campos de estudio específicos (Cross et al., 1981). Este conocimiento se organiza en función de principios explicativos, ya sea de forma teórica o práctica. Al emplear el razonamiento lógico y el análisis objetivo de la evidencia científica, se formulan investigaciones e hipótesis, se deducen principios y leyes y se construyen modelos, teorías y sistemas de conocimiento utilizando el método científico (Skinner, 1956). La ciencia reconoce y se basa en la observación experimental. Dicha observación se estructura a través de métodos, modelos, y teorías para generar conocimientos novedosos. Para lograr este objetivo, se establecen de antemano criterios de verdad y un enfoque investigativo. La aplicación de estos enfoques y conocimientos da como resultado la creación de nuevos conocimientos en forma de predicciones concretas, cuantitativas, y comprobables sobre las observaciones pasadas, presentes y futuras (Carey & Carey, 1994). Con frecuencia, estas predicciones pueden formularse mediante el razonamiento y organizarse como reglas o leyes generales que explican el comportamiento de un sistema y predicen sus acciones en circunstancias particulares. Sin embargo, esta forma de conocer el mundo sistemática, verificable, y orientada a la predicción no siempre fue la norma. Su consolidación como método dominante fue el resultado de tres transformaciones históricas fundamentales que redefinieron la relación del ser humano con el conocimiento, la moral, y la política.

Tres transformaciones fundamentales que caracterizaron el cambio del mundo medieval convencional al contemporáneo son de suma importancia. Estas incluyen el surgimiento de las ciencias experimentales y la adopción del novedoso método de conocimiento

científico; la crítica racional de todos los elementos heredados, que abarca la religión y las diversas facetas de una realidad fragmentada; y las alteraciones en los sistemas éticos y políticos que otorgan al individuo autonomía moral y redefinen su relación con la comunidad. Este análisis examina las diversas trayectorias que se originan en el legado griego y helenístico, así como en el ámbito judeocristiano, los cuales son hitos clave en la cultura occidental y se extienden a las figuras prominentes que desempeñaron un papel crucial en la configuración del mundo moderno, como Maquiavelo, Montaigne, y Galileo (Popper, 1959; Vázquez, 2021). Estas trayectorias redefinen la comprensión de la realidad y la comprensión de la naturaleza y las interacciones humanas, impulsando así un cambio de paradigma. Este proceso transformador tuvo un profundo impacto en todos los ámbitos, incluidos el cosmos, el mundo viviente y la esfera política. Mientras Copérnico y Galileo dismantelaban la estructura jerárquica del universo, Montaigne desafió las disparidades psicológicas entre los individuos y Maquiavelo revolucionó la percepción del ámbito político (Vázquez, 2021).

Durante el Renacimiento, el entusiasmo inicial por los clásicos de la ciencia griega fue seguido por una profunda sensación de desilusión a la luz de sus falacias, junto con una mayor valoración del conocimiento experiencial personal (Skinner, 1956). La llegada de la imprenta facilitó la formación de una alianza sin precedentes entre artesanos e intelectuales selectos con un gran interés en el estudio del mundo natural. El descubrimiento del Nuevo Mundo supuso una vía innovadora para evaluar las propias experiencias a través de los viajes y la observación del cosmos. El período barroco fue testigo del surgimiento del método experimental, que podía probarse empíricamente debido a la disponibilidad de herramientas novedosas para la observación y la medición. Las matemáticas introdujeron nuevos recursos analíticos. En el campo de la astronomía, la noción de la Tierra como un simple cuerpo celeste y la aceptación del heliocentrismo se afianzaron firmemente. La institucionalización de la investigación científica echó raíces en Inglaterra y Francia (Skinner, 1956). Poco a poco, la ciencia emergió como un paradigma de pensamiento racional y lógico, basado en la investigación individual, que también generó transformaciones evidentes en la actividad económica y la estructura social a través de los avances tecnológicos y cimentó los orígenes del método científico (Vázquez, 2021).

Actualmente, el método científico se distingue por el enfoque empleado más que por el tema de investigación. Se refiere a una serie de procedimientos dispuestos para

comprender la realidad (Popper, 1959; Privitera, 2022). Ver Figura 1 sobre el proceso del método científico. El procedimiento principal que debe ejecutarse es la investigación de la realidad para lograr una comprensión imparcial, lo que implica que no puede tratarse de ningún tipo de indagación. Los requisitos previos para llevar a cabo una investigación adecuado son la organización, la regulación y la cuantificación. En consecuencia, podemos definir el método científico como el conjunto de procedimientos organizados, mediante los cuales el objetivo es obtener controles metódicos y regulados, que posteriormente se cuantificarán y documentarán, lo que nos permitirá obtener una comprensión imparcial del fenómeno que se está examinando.

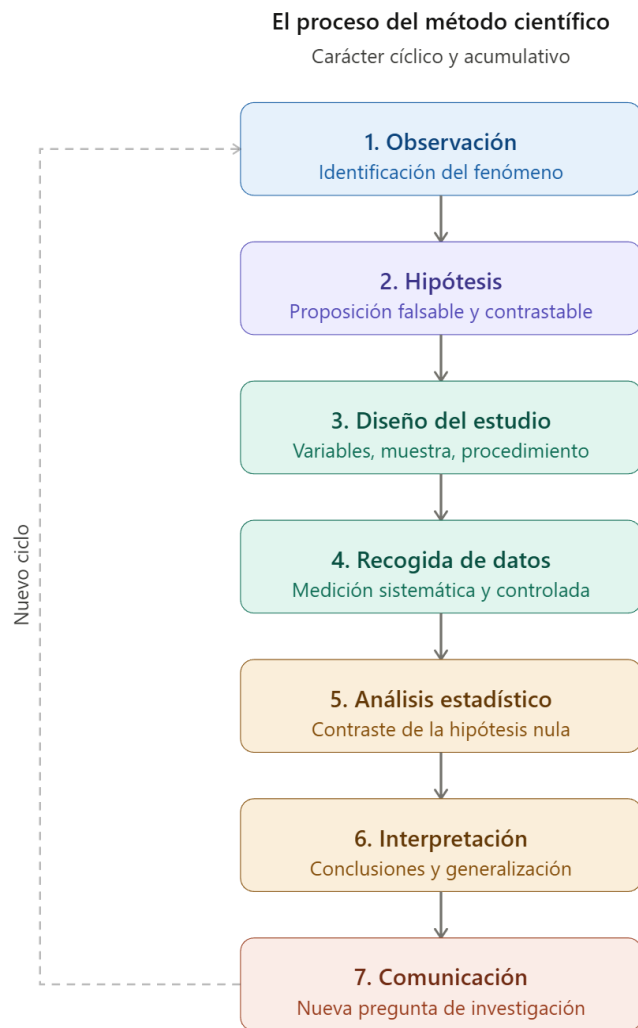
Método científico: conjunto de procedimientos sistemáticos y replicables mediante los cuales se genera y valida el conocimiento objetivo. Su rasgo definitorio es la contrastación empírica: toda afirmación debe poder someterse a prueba y ser potencialmente falsable (Popper, 1959).

Durante más de un siglo, la Psicología ha puesto en práctica el método científico como base para su investigación. El método científico se refiere a la forma en que los científicos se plantean preguntas y a la lógica y los métodos utilizados para llegar a esas respuestas. Existen diversas perspectivas para adquirir conocimientos sobre nosotros mismos y nuestro entorno, como la filosofía, la teología, la literatura, el arte y otras disciplinas (Cross et al., 1981). Es por ello por lo que el método científico se distingue de otros métodos, pero todos comparten el mismo fin: la búsqueda de la verdad.

Uno de los enfoques para comprender el método científico para la búsqueda de la verdad es distinguirlo de las concepciones e inferencias que utilizamos para adquirir nuestro conocimiento. Al igual que el telescopio y el microscopio amplían nuestra capacidad de ver, el método científico amplía nuestra forma cotidiana de saber (Carey & Carey, 1994). Los científicos que se dedican a la Psicología son cautos a la hora de aceptar aseveraciones acerca del comportamiento y los mecanismos mentales, y examinan críticamente las pruebas antes de validar dichas aseveraciones. No obstante, nuestro razonamiento cotidiano nos lleva a aceptar pruebas y aserciones con poca o ninguna evaluación de las pruebas. En general, hacemos muchos de nuestros juicios cotidianos utilizando la intuición. Esto suele significar que actuamos basándonos en lo que nos parece bien o lo que parece razonable. Cuando nos basamos en la intuición para emitir juicios, a menudo no reconocemos que nuestras percepciones pueden estar distorsionadas

por sesgos cognitivos o que puede que no hayamos tenido en cuenta todas las pruebas disponibles (Tversky & Kahneman, 1973, 1974). El método científico que usamos para lograr conocimiento tiene un enfoque empirista. Dicho enfoque hace hincapié en la metodología observacional y la realización de experimentos para responder a las hipótesis planteadas (Skinner, 1956). La intuición desempeña un papel esencial en este proceso, y es frecuentemente el origen de las preguntas de investigación, la chispa que lleva al científico a formular una hipótesis novedosa o a percibir una relación que los datos aún no han confirmado. Sin ella, la ciencia difícilmente generaría ideas originales. Sin embargo, la intuición por sí sola no es suficiente. Una vez formulada la pregunta, el científico se apoya en las pruebas empíricas que proporcionan la observación directa y la experimentación para determinar si esa intuición se sostiene o debe ser revisada. Es precisamente esta combinación, la intuición como motor generador de hipótesis y la evidencia empírica como criterio de validación, lo que define el carácter científico de la Psicología como disciplina, y lo que hace necesario conocer con precisión los métodos a través de los cuales dicha evidencia se genera. En el apartado siguiente se examina cómo se ha articulado la investigación en Psicología en torno a ese propósito.

Figura 1: El proceso del método científico (*flowchart* cíclico).



Fuente: adaptado de Privitera (2022) y Shadish et al. (2002)

En resumen

- El método científico es el conjunto de procedimientos sistemáticos mediante los que la ciencia genera y valida conocimiento. Su rasgo definitorio es la contrastación empírica: toda afirmación debe poder someterse a prueba.
- Se distingue de otras formas de conocimiento (filosofía, intuición, sentido común) en que no acepta ninguna afirmación sin evidencia contrastable empíricamente.

- Su aplicación en Psicología sigue un ciclo: observación del fenómeno → formulación de hipótesis → diseño del estudio → recogida y análisis de datos → interpretación → nueva pregunta.
- La Psicología utiliza este método desde hace más de un siglo para estudiar el comportamiento y los procesos mentales con el mismo rigor que las ciencias naturales.

2. La investigación en la Psicología

Definiciones:

Variable independiente (VI): factor que el investigador manipula de forma deliberada para estudiar su efecto sobre el fenómeno de interés.

Variable dependiente (VD): factor que el investigador mide para detectar el efecto de la manipulación; es el criterio de resultado del estudio.

Variable extraña (VE): cualquier factor distinto de la VI que puede influir sobre la VD y que, si no se controla, amenaza la validez de las conclusiones.

Hipótesis: proposición empíricamente contrastable que establece una relación esperada entre la VI y la VD. En investigación experimental adopta la forma: «La manipulación de X producirá cambios en Y bajo las condiciones Z».

Si el método científico es el marco general, la investigación en Psicología es su aplicación concreta al estudio del comportamiento humano. Antes de diseñar cualquier estudio, el investigador necesita manejar con precisión tres conceptos fundamentales: la variable que va a manipular, la que va a medir y las que puede que interfieran sin que lo haya planeado. Comprender estos tres elementos es el punto de partida de cualquier metodología de investigación. El método científico constituye el marco general que guía la búsqueda del conocimiento, la investigación en Psicología ha desarrollado formas propias de aplicarlo al estudio del comportamiento y los procesos mentales. Comprender en qué consiste investigar en Psicología, qué se manipula, qué se mide y cómo se organizan los distintos elementos del proceso, es el punto de partida indispensable para entender cualquier metodología concreta. El enfoque experimental abarca el examen, la manipulación y la documentación de los diversos factores (como las variables dependientes, independientes e intermedias) que afectan a un tema de investigación (Vautier, 2011). En el contexto específico de la Psicología, es posible delinear y dilucidar estos factores en relación con la conducta humana, lo que permitiría predecir sus alteraciones (Haig, 2014). Los psicólogos experimentales tienen un gran interés en comprender el comportamiento individual y emplean medidas controladas en el laboratorio para influir o impactar potencialmente en la vida. Emplean instrumentos precisos y necesitan un alto nivel de regulación y cuantificación. El término «Psicología Experimental» denota principalmente

una clasificación de la Psicología basada en criterios metodológicos más que en consideraciones sustantivas (Pascual et al., 1995). En consecuencia, cualquier escuela o corriente psicológica que adopte el método experimental se clasifica como parte de la Psicología Experimental (Kirk, 2013), independientemente de las perspectivas epistemológicas relacionadas con su tema de investigación. Dominios como la percepción, la memoria, el aprendizaje y la cognición se han investigado tradicionalmente mediante el uso del enfoque experimental (Balluerka & Vergara, 2002; Pascual et al., 1991).

El modelo general de investigación en el campo de la Psicología, situado en el ámbito de las ciencias sociales y psicológicas, se organiza en tres niveles distintos, organizados formalmente de manera jerárquica como nivel de expectativa teórica, nivel de diseño y nivel de análisis de datos (Anguera et al., 1995). Según este modelo general, el nivel inicial establece la esencia conceptual del modelo. En esta etapa, se determinan los límites de la experiencia observada, se formulan preguntas o cuestiones en función de las observaciones realizadas, se presentan posibles explicaciones teóricas en forma de hipótesis o conjeturas, y de estas hipótesis se obtienen resultados comprobables empíricamente. Estos resultados sirven como hipótesis sobre las que se planifica el diseño, dependiendo de un enfoque específico de recopilación de datos. De esta manera, el nivel de expectativas teóricas se centra principalmente en las hipótesis o expectativas sobre la forma en que las variables están interconectadas, así como en las posibles asociaciones entre los fenómenos observados que definen la base empírica de una disciplina científica.

Ejemplo: Un investigador observa que los estudiantes que duermen menos de seis horas rinden peor en los exámenes. A partir de esa observación formula la hipótesis de que la privación de sueño reduce el rendimiento en memoria de trabajo. Esta hipótesis, todavía en el plano teórico, delimita el fenómeno de interés, anticipa una relación entre dos variables y genera una predicción comprobable empíricamente: si se reduce el sueño, el rendimiento en memoria disminuirá.

En un nivel posterior, caracterizado por la operacionalización de las hipótesis, se delinean las estrategias que se emplearán para adquirir la información necesaria y abordar los desafíos teóricos planteados. La tarea principal de este segundo nivel radica en la identificación técnica y práctica de las variables que constituyen el escenario de

investigación. Para cada una de estas variables, el investigador determina el alcance de la intervención, es decir, si se manipula, se mide o simplemente se controla mediante diversas técnicas. El objetivo de este segundo nivel es adquirir información relacionada con las diferentes variables involucradas en el diseño del estudio o proyecto.

Ejemplo: Continuando con el caso anterior, el investigador debe ahora concretar la hipótesis en decisiones técnicas precisas. La privación de sueño (4 horas vs. 8 horas) es la variable que decide manipular, él mismo asigna a cada participante a una de las dos condiciones. El rendimiento en memoria de trabajo (puntuación en una prueba estandarizada de recuerdo de dígitos) es la variable que decide medir, es el resultado que observará para comprobar si la manipulación ha tenido efecto. En nivel previo de vocabulario, la edad y el consumo de cafeína son variables que no manipula ni mide como objetivo principal, pero que podrían distorsionar los resultados si no se igualan entre grupos: son variables que decide controlar. Este proceso, decidir qué se manipula, qué se mide y qué se controla, es la operacionalización, el puente entre la pregunta teórica y el diseño práctico del estudio.

A nivel de análisis, el investigador emplea modelos estadísticos y las correspondientes pruebas de significación. Para ello, inicialmente se postula la hipótesis nula con el objetivo de refutarla. Dependiendo del resultado de la prueba estadística, los datos se interpretan a la luz de los supuestos y modelos teóricos que se han empleado a lo largo del proceso de investigación. Es importante señalar que el modelo de análisis depende de dos consideraciones fundamentales: la naturaleza de los datos, ya sean cuantitativos o categóricos, y la estructura del diseño, incluida la cantidad de grupos o variables independientes.

Ejemplo: Con los datos recogidos, el investigador comprueba si la diferencia en rendimiento entre el grupo de 4 horas y el de 8 horas es estadísticamente significativa. Para ello parte de la hipótesis nula, no existe diferencia en el rendimiento entre ambas condiciones, e intenta refutarla mediante una prueba t para muestras independientes, dado que los datos son cuantitativos (puntuaciones de memoria) y el diseño compara dos grupos. Si el resultado permite rechazar la hipótesis nula ($p < .05$), el investigador concluye que la privación de sueño afecta significativamente al rendimiento en memoria de trabajo, e interpreta ese hallazgo a la luz de los modelos teóricos sobre sueño y cognición revisados en el primer nivel.

De toda la información que se ha proporcionado hasta ahora, se puede deducir que el modelo de investigación global sigue un proceso secuencial que se organiza en diferentes niveles. Cada uno de estos niveles se guía por un conjunto de suposiciones y regulaciones relacionadas con el papel del desempeño dentro del proceso general. A su vez, las metodologías de trabajo se refieren a procedimientos específicos que se emplean para abordar los problemas de investigación y sirven como mecanismo de conexión entre el modelo de investigación general y una situación de estudio particular. Si bien la ciencia psicológica ofrece una amplia gama de metodologías, se pueden distinguir en función del nivel de rigor y poder inferencial que poseen, es decir, la medida en que controlan las causas o variables que influyen en la variación de la variable dependiente u objeto de análisis (Balluerka & Vergara, 2002).

En resumen

- La investigación en Psicología se organiza en tres niveles: expectativa teórica (hipótesis), diseño (cómo recoger los datos) y análisis (cómo interpretarlos).
- Tres conceptos vertebran cualquier estudio: variable independiente (VI, lo que manipulamos), variable dependiente (VD, lo que medimos) e hipótesis (la predicción sobre la relación entre ambas).
- Las variables extrañas (VE) son factores ajenos al estudio que pueden distorsionar los resultados si no se controlan: su gestión es uno de los pilares del diseño experimental.
- El nivel de control sobre las variables determina el poder inferencial del estudio: cuanto mayor el control, más confianza tenemos en que los resultados son atribuibles a la VI.

3. Metodología de Investigación en Psicología

La investigación realizada en el campo de la Psicología desempeña un papel crucial en el avance de la disciplina como ciencia. En el ámbito de la Psicología, existe una rica tradición de investigación que abarca varias clasificaciones y nomenclaturas, que dependen de la perspectiva del autor y del contexto específico en el que se lleva a cabo la investigación. Los fundamentos ontológicos (¿qué existe?), gnoseológicos (¿podemos conocerlo?) y epistemológicos (¿cómo lo conocemos científicamente?) sirven de base para los métodos empleados en la investigación psicológica, ya que estos fundamentos tienen implicaciones tanto académicas como prácticas (Haig, 2014). Esta noción se refiere al establecimiento de definiciones terminológicas, específicamente, a los conceptos de metodología, método y técnica (Valsiner et al., 2017). La metodología abarca un examen reflexivo de los procedimientos científicos y los instrumentos de investigación empleados, que implica el análisis crítico y la interrelación de los diferentes métodos en relación con los estándares comunitarios. Básicamente, la metodología describe los pasos necesarios que deben tomarse para que una actividad en particular sea reconocida como científica (Gómez-Luna et al., 2014). Si bien la disciplina metodológica posee una dimensión descriptiva que comparte con la historia de la ciencia, las instituciones educativas contemporáneas enfatizan su naturaleza prescriptiva y empírica. Esto implica el examen de las experiencias de varios investigadores y la evaluación de sus caminos de descubrimiento, todo ello sin ningún tipo de nociones preconcebidas.

Además, la metodología en sí misma, cuando se refiere a las técnicas empleadas en la implementación de un diseño de investigación, se refiere a la utilización de métodos para adquirir nuevos conocimientos y para validar los hallazgos obtenidos (Gauch, 2015). El primer caso se refiere a la metodología, mientras que el segundo se adentra en el ámbito de la epistemología. Los científicos se basan en métodos, que posteriormente son validados por los epistemólogos, y estos métodos son diseñados por metodólogos. Hernández et al., (2006) sostienen que el campo de la Psicología se muestra una inclinación significativa hacia el pluralismo metodológico y la toma de decisiones autónoma de los investigadores. Según su perspectiva, el avance científico responde a un principio de acumulación progresiva: a mayor volumen de investigación rigurosa, mayor es el crecimiento del conocimiento disponible en un campo, aunque ese crecimiento no sea lineal ni automático. En el ámbito de la investigación científica, los principios de organización, rigor y atención son de suma importancia. Sin embargo, es crucial

reconocer que este proceso de investigación no es inflexible, ya que la investigación científica es intrínsecamente de naturaleza dinámica, fluida y evolutiva. Los estudiosos sostienen que la investigación científica incorpora una fusión de procedimientos sistemáticos y empíricos que están diseñados específicamente para investigar y comprender fenómenos complejos (Cross et al., 1981; Gómez-Luna et al., 2014).

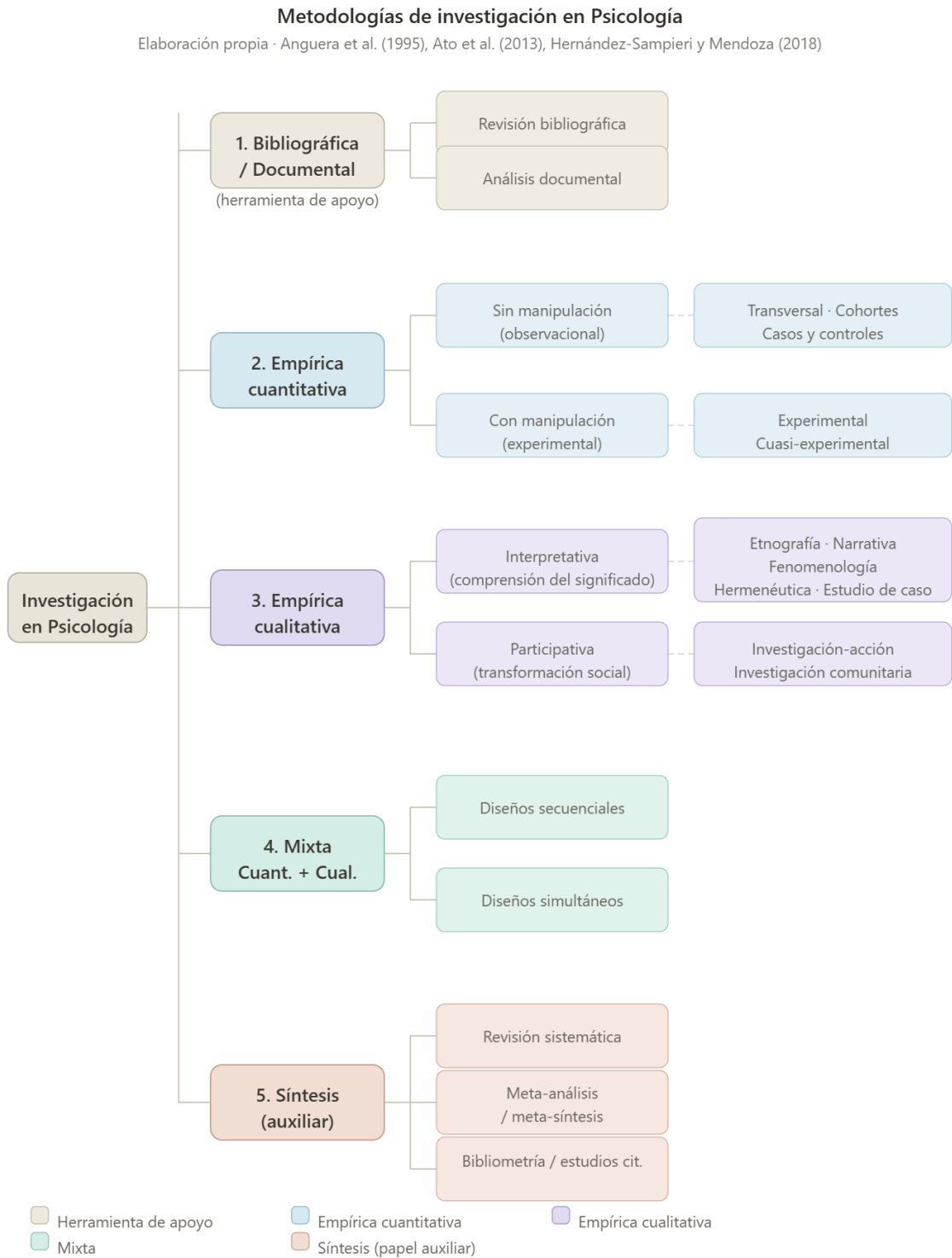
En Psicología existen diferentes metodologías de investigación y existen distintas estrategias para realizar la clasificación de las metodologías empleadas. Quizás uno de los aspectos más importantes a tener en consideración en las distintas clasificaciones es el error de no tener un método definido y no saber cómo aplicarlo (Simmons et al., 2011; Pascual et al., 1995). Una de las clasificaciones más clásicas es la diferenciación entre las metodologías cuantitativas y cualitativas (Arnau, 1975; Campbell & Stanley, 1979). Por su lado, Hernández-Sampieri y Mendoza (2018) agrupan coincidentemente los estudios cualitativos y cuantitativos, pero complementa el campo de investigación mixta (Creswell & Creswell, 2018). Desde el punto de vista metodológico, la propuesta de Arnau (1990) distribuye el paradigma metodológico cuantitativo en tres grandes categorías: descriptivo (que incluye los diseños observacionales y de encuesta), explicativo (orientado a establecer relaciones entre variables) y experimental (que engloba los diseños experimentales y cuasi-experimentales). Otras clasificaciones han distinguido entre la metodología experimental o manipulativa y no-experimental, no-manipulativa o descriptiva (Arnau, 1989; León & Montero, 2004). Por otro lado, otros autores utilizan la clasificación de estrategias manipulativa entre metodología experimental y cuasi-experimental (Anguera, 2016; Ato et al., 2013; León & Montero, 2004), y en la estrategia no-manipulativa entre observacional y selectiva o de encuesta, o entre observacional y correlacional (Arnau et al., 1990; Campbell & Stanley, 1979).

Los postulados que se han considerado en esta revisión se refieren a la clasificación de las perspectivas metodológicas dentro del campo de la Psicología. Sin embargo, es importante señalar que estos postulados no tienen en cuenta la inclusión de las revisiones y los meta-análisis como una entidad independiente dentro del marco teórico. Se argumenta que al reconocer el valor y el progreso que se pueden lograr a través de estas representaciones, el campo de la Psicología puede avanzar como ciencia. Esto se debe a que estas representaciones permiten la combinación y comparación de los resultados de la investigación (Marín et al., 2009). Como puede observarse, la literatura metodológica ha propuesto diversas formas de clasificar los métodos de investigación en Psicología.

Las dos más ampliamente aceptadas son: (a) la distinción entre metodologías con estrategia manipulativa, fundamentalmente la metodología experimental, y metodologías sin estrategia manipulativa, observacional, correlacional y selectiva, (Arnau, 1989; León & Montero, 2004); y (b) la tripartición clásica en metodología observacional, metodología correlacional o selectiva y metodología experimental (Anguera et al., 1995; Ato et al., 2013). La propuesta que se adopta en este manual no pretende sustituir estas taxonomías consolidadas, sino complementarlas para reflejar el amplio panorama metodológico actual de la Psicología, en el que coexisten enfoques empíricos cuantitativos y cualitativos, metodologías mixtas y metodologías de síntesis. Dado que el estudiante de grado en Psicología ha de familiarizarse con este conjunto amplio de estrategias de investigación antes de especializarse en alguna de ellas, se opta por una clasificación inclusiva de cinco subgrupos que facilite esa orientación inicial.

A la luz de esto, se presenta la siguiente propuesta como un medio para clasificar las diversas metodologías utilizadas en Psicología. El sistema de clasificación propuesto tiene como objetivo sistematizar las opciones de investigación actuales disponibles en el campo de la Psicología. Al hacerlo, proporciona un marco inclusivo y abierto que abarca el paradigma de la complejidad. Además, este sistema de clasificación propuesto permite el progreso disciplinario al incorporar recursos que ya se utilizan en la investigación psicológica. Como resultado, el sistema de clasificación propuesto clasifica los métodos de investigación en Psicología en cinco subgrupos. El primer subgrupo consiste en metodologías utilizadas en los análisis bibliográficos y/o documentales para proporcionar una perspectiva específica sobre un tema en particular dentro de la ciencia psicológica. Estas metodologías se caracterizan por su contenido teórico y conceptual. El segundo subgrupo abarca las metodologías empíricas cuantitativas (tanto descriptivos como explicativos), mientras que el tercer subgrupo se centra en las metodologías empíricas cualitativas. El cuarto subgrupo incluye metodologías mixtas que combinan paradigmas de investigación cualitativos y cuantitativos. Por último, el quinto subgrupo está dedicado a las metodologías de revisión: las revisiones narrativas, las revisiones sistemáticas, la meta-síntesis, el meta-análisis y los estudios bibliométricos. Ver Figura 2.

Figura 2: Mapa de las cinco categorías metodológicas de la Sección 3.



3.1. Metodologías de análisis bibliográfico y/o documental

Las metodologías de análisis bibliográfico y documental no constituyen, en sentido estricto, metodologías de investigación psicológica autónomas, en ellas el documento es

el objeto de análisis, no el comportamiento o los procesos psicológicos, sino herramientas de revisión y síntesis que sirven de apoyo a cualquier proyecto de investigación empírica. Su inclusión en este manual se justifica por su utilidad práctica para el estudiante de grado, que desde el inicio de su formación necesita dominar las estrategias de búsqueda, selección y gestión de fuentes bibliográficas. La metodología documental implica el escrutinio de fuentes primarias, esto es, datos e información escrita que no han sido objeto de tratamiento previo (Sánchez & Valdés, 2008); la investigación bibliográfica, por su parte, se apoya en fuentes secundarias como libros, artículos, disertaciones y tesis localizados en bibliotecas y repositorios (Gómez-Luna et al., 2014). Estas dos metodologías se complementan entre sí, proporcionando una visión integral del tema objeto de investigación en el ámbito de la investigación cualitativa. Esta metodología se inscribe en el dominio de la investigación cualitativa y tiene sus raíces en la investigación bibliográfica o de revisión (Perin & Pita, 2019). En ella se proporciona la descripción detallada de un tema específicos, sin considerar la identificación de tendencias que puedan presentar diferentes escenarios para el desarrollo de dicha tecnología y permitir tomar decisiones. La metodología descrita para la revisión bibliográfica consta de tres fases esenciales, que deben incorporarse de forma continua durante todo el proceso de búsqueda. La metodología propuesta abarca varias fases, como es la definición del problema. Donde se clarifica de forma exhaustiva la búsqueda bibliográfica, para que cumpla con los requisitos del investigador específico y, al mismo tiempo, contribuya al estado actual de los conocimientos (Sánchez & Valdés, 2008). Como resultado, se establece un escenario amplio que garantiza que la investigación pueda tener una recuperación. Para el proceso de recuperación de información en la investigación bibliográfica, deben estar accesibles los materiales de información relevantes, como libros, revistas de investigación populares o académicas, sitios web y otras fuentes necesarias. Un enfoque estructurado y profesional es esencial para realizar una búsqueda bibliográfica, ya que leer la documentación sin rumbo fijo es un ejercicio inútil. Naturalmente, al principio de una búsqueda bibliográfica, es imposible discernir qué materiales son los más pertinentes o relevantes. No obstante, en el avance de la investigación, la vista completa se hace más clara y el investigador comienza a identificar los temas que realmente despiertan su interés. Por lo tanto, es crucial delimitar la búsqueda y saber cuándo concluir, incluso si aún quedan preguntas sin respuesta antes de abordar el enfoque principal del proyecto (Gómez-Luna et al., 2014).

La organización de la información tiene una importancia significativa en cualquier proyecto de investigación, ya que implica la organización sistemática de la documentación recopilada. Esto se puede hacer de forma manual, pero este proceso es lento e ineficaz. Alternativamente, se pueden utilizar una serie de paquetes informáticos que facilitan la tarea (JabRef, Zotero, Mendeley, Endnote y Reference Manager) y permite una organización más eficiente. El análisis de la información representa la tercera fase, en la que se examina la información ya organizada para determinar los documentos más útiles para el tema de estudio. Esta tarea de análisis de la información es el aspecto que consume más tiempo de la investigación bibliográfica, ya que tiene como objetivo identificar la contribución que se puede hacer. El pensamiento crítico debe emplearse a lo largo de esta fase en paralelo con la fase inicial, ya que es un proceso continuo. Es un proceso cíclico en el que se reafirman las ideas formuladas en el planteamiento del problema y, si el problema se entiende bien, la solución se vuelve alcanzable. Numerosos estudios han aplicado con éxito diversas estrategias para un análisis exhaustivo de la información, con resultados satisfactorios (Gómez-Luna et al., 2014).

3.2. Metodologías empíricas cuantitativas

La investigación cuantitativa se refiere a la recopilación y el análisis de datos cuantitativos sobre variables y al estudio de propiedades y fenómenos cuantitativos. Se emplean varias técnicas de análisis, como el análisis descriptivo, el análisis exploratorio, la inferencia univariante, la inferencia multivariante, la modelización y el contraste (Creswell & Creswell, 2018; Sampieri & Mendoza, 2018). Estos estudios pueden ser longitudinales, en los que se observa a la población estudiada durante un período de tiempo, o no longitudinales, sin seguimiento temporal. En los estudios prospectivos, el efecto se produce después del inicio del estudio, mientras que, en los estudios retrospectivos, el efecto ya se produjo cuando comenzó el estudio. El enfoque cuantitativo se caracteriza por una serie de pasos que sigue el investigador (Anguera & Hernández-Mendo, 2016). En primer lugar, se plantea un problema de estudio específico y concreto. Luego, a partir de una revisión de la literatura, se construye un marco teórico. Las hipótesis se derivan de esta teoría y se utilizan diseños de investigación apropiados para probar estas hipótesis. Si los resultados apoyan o se alinean con las hipótesis, se proporcionan pruebas que respalden dichas hipótesis. Para lograr estos resultados, el investigador recopila datos numéricos de objetos, fenómenos o participantes del estudio y aplica procedimientos estadísticos para el análisis (Arnau et al., 1990). Las hipótesis se formulan antes de la

recopilación y el análisis de los datos, lo que subraya la importancia de establecer hipótesis con antelación. La recopilación de datos implica la medición, en la que se miden las variables o conceptos descritos en las hipótesis. A medida que los datos son el resultado de las mediciones, se representan cuantitativamente y es necesario analizarlos mediante enfoques estadísticos. El objetivo principal es garantizar el máximo control y mitigar las explicaciones alternativas de la propuesta de estudio (hipótesis), reduciendo así la incertidumbre y el error. Los análisis cuantitativos se interpretan a la luz de las hipótesis iniciales y de los estudios de campo anteriores. Es fundamental que la investigación cuantitativa mantenga la objetividad en todo momento. Estos estudios siguen una secuencia previsible y estructurada, y las decisiones fundamentales se toman antes de la recopilación de datos. La esencia de la investigación cuantitativa es dilucidar y pronosticar los fenómenos investigados, buscando patrones y relaciones causales entre los elementos. Los datos producidos se adhieren a criterios de validez y confiabilidad, y las conclusiones derivadas contribuyen a la generación de conocimiento. El enfoque cuantitativo emplea un razonamiento lógico o deductivo, partiendo de la teoría y derivando expresiones lógicas conocidas como hipótesis, que están destinadas a ser probadas por el investigador. En última instancia, la investigación cuantitativa gira en torno a la realidad externa del individuo (Bandalos & Finney, 2018; Thurstone, 1947).

Desde la perspectiva de la existencia de la manipulación de las variables en la metodología empírica cuantitativa, se pueden realizar la clasificación en estudios sin manipulación de variables, llamados estudios de observación, y los estudios que conllevan alguna manipulación de variables, conocidos como estudios de intervención (Campbell & Stanley, 1979). Incluidos en la clasificación de estudios sin manipulación se pueden encontrar estudios transversales, y estudios de casos y controles. Por otro lado, en los estudios con manipulación de variables se encuentran los estudios experimentales y cuasi-experimentales.

Los estudios transversales representan una forma fundamental de diseño de investigación. Emplean un enfoque transversal no experimental, en el que se examina una comunidad específica o una muestra representativa en un momento determinado. La evaluación de las variables se produce simultáneamente. Es esencial garantizar que la muestra seleccionada represente con precisión a la población del estudio. Cada sujeto del estudio solo se evalúa una vez. Los estudios descriptivos tienen como objetivo describir las variables dentro de un grupo de sujetos durante un período breve, sin incorporar un grupo

de control (Bandalos & Finney, 2018). Estos estudios son relativamente rápidos y rentables, y sirven como paso inicial para estudios posteriores más complejos. Sin embargo, no establecen una secuencia de eventos ni relaciones causales, no determinan la incidencia o el riesgo relativo, y pueden conllevar posibles sesgos (Bakeman & Queram 2023). Sin embargo, los estudios de cohortes son diseños no experimentales orientados hacia el futuro, basados en el conocimiento de la exposición a una situación o condición. Los estudios de cohorte abarcan dos cohortes: una expuesta al fenómeno objeto de investigación y otra no expuesta, seleccionadas de forma independiente. Un estudio de cohorte constituye un diseño analítico observacional en el que los participantes se organizan según la exposición hipotética a un factor de riesgo para una enfermedad o un efecto. Posteriormente, se hace un seguimiento de estas personas durante un período de tiempo para evaluar la incidencia del resultado.

Los estudios de casos y controles se definen como diseños analíticos epidemiológicos no experimentales con una orientación retrospectiva, que comienzan con la investigación de la historia del efecto. Estos estudios implican la selección de dos grupos de sujetos: un grupo de control que carece de la enfermedad o el efecto objeto de examen, y un grupo de casos sin un examen previo de la enfermedad o el efecto. Estos grupos se comparan en función de las exposiciones o las características de fondo para determinar su asociación con el efecto estudiado (González-Garay et al., 2018). En resumen, un estudio de casos y controles constituye un diseño analítico observacional en el que los sujetos se eligen en función de la existencia de un efecto o circunstancia. En tal caso se debe procurar la coherencia, con criterios deben estar claramente definidos. Los controles deben abarcar a todos los sujetos que hayan tenido la oportunidad de exponerse, excluyendo a los que no puedan exponerse por cualquier motivo. Además, los controles deben representar con precisión a la población de la que se derivan los casos.

Dentro de los métodos cuantitativos con manipulación de variables, la metodología experimental tiene como objetivo determinar cuál es el impacto que tiene una variable considerada como independiente sobre otra variable considerada como dependiente (Arnau et al., 1990; Campbell & Stanley, 1979). Por lo tanto, el objetivo principal de la metodología experimental es estudiar los cambios que se producen en la variable dependiente (VD) como resultado de las modificaciones de la variable independiente (VI), manteniendo al mismo tiempo un control estricto sobre cualquier variable contaminante o extraña (VE) (Bandalos & Finney, 2018). En este contexto, los cambios

observados en la VD como consecuencia de las variaciones en la VI pueden establecer una relación causal entre ambas. Los requisitos esenciales para clasificar la investigación como experimental son la manipulación intencional de la VI, que implica una manipulación activa y una variación controlada. Además, la VI debe tener al menos dos niveles o condiciones para facilitar la comparación. Por otro lado, es necesaria la observación sistemática del efecto sobre la VD, que implica una medición objetiva para detectar cómo influye en ella la manipulación de la VI. El último de los requisitos de la investigación experimental es el control de las VE, que incluye todas las variables distintas de la VI que podrían influir en la medición de la VD. Estas variables pueden provenir de diversas fuentes, como el sujeto experimental, el experimentador, el contexto y el entorno, así como de la medición de la VD y la manipulación de la VI (Anguera & Hernández-Mendo, 2016). El control de las variables, así como el grado de generalización de los resultados determina lo que se conoce como validez de la investigación, desarrollado en el punto 5. Hay varias técnicas de control disponibles según la fuente de las variables contaminantes. Para las que provienen del sujeto experimental, las técnicas de control más comunes incluyen la aleatorización (la más utilizada), la comparación, el control estadístico, el procedimiento ciego simple y el sujeto como control propio. Para las variables contaminantes procedentes del experimentador, se emplean procedimientos de equilibrio y doble ciego. Las técnicas de control para las variables contaminantes del contexto y el medio ambiente incluyen la eliminación, la constancia y el balanceo. Por último, las técnicas de control de las variables contaminantes derivadas de la medición de la VD y la manipulación de la VI incluyen la asignación al azar de los niveles de la VI a los grupos experimentales, la aleatorización del orden de aplicación de los niveles de la VI a los sujetos y el contrabalanceo (Arnau, 1975; Campbell & Stanley, 1979).

El aspecto fundamental de la metodología experimental radica en la utilización de los experimentos como situaciones artificiales en las que se establecen las condiciones necesarias para aplicar diferentes valores a la VI y evaluar los efectos resultantes sobre la VD. Arnau (1978) ha definido este concepto como la representación de un escenario real que el investigador crea artificialmente para investigar el comportamiento de las variables y su significado funcional. Dentro de estas situaciones artificiales, se pueden identificar dos tipos de experimentos. En primer lugar, están los experimentos de laboratorio que se llevan a cabo en condiciones de total artificialidad, en los que se construyen artificialmente las condiciones necesarias para la aparición de variables independientes.

En consecuencia, el contexto en el que se lleva a cabo la investigación es muy poco representativo, con una validez externa y ecológica limitadas. Por el contrario, los experimentos de campo tienen lugar en entornos naturales, donde el investigador manipula y controla las variables de interés. Esto se traduce en una mayor validez externa y, en particular, en una validez ecológica de la investigación. Sin embargo, el control de variables ajenas se vuelve más complejo en estos experimentos.

La metodología cuasi-experimental se diferencia de la metodología experimental en el grado de control que mantiene el experimentador sobre la variable manipulada ya que en este caso no es total y por lo tanto puede haber contextos del desarrollo o de procedimientos que estén afectando sin que estos sean controlados (Cook & Campbell, 1979). Cuando el experimentador carece de un control total sobre las situaciones que se están investigando, no podemos referirnos a ello como un experimento genuino, sino más bien como cuasi-experimentos. En tales casos, las investigaciones deberían etiquetarse como cuasi-experimentales. Las características clave de la metodología cuasi-experimental son que el experimentador no puede lograr un control experimental completo. Además, estos experimentos se llevan a cabo en contextos naturales, a menudo de naturaleza social. Se implementan estrategias metodológicas, como la realización de observaciones múltiples, para mitigar la ausencia de control experimental y las posibles amenazas a la validez interna y la investigación. Estas estrategias se emplean principalmente cuando no es posible llevar a cabo un estudio de investigación experimental exhaustivo (Campbell & Stanley, 1979).

En el campo aplicado, la investigación cuasi-experimental surge como resultado de una serie de circunstancias. Una de esas circunstancias es cuando la VI solo puede manipularse mediante una selección de valores predeterminados. Este escenario surge con frecuencia en los campos de la Educación y la Psicología cuando se examina la relación entre la VI del sujeto y una variable dependiente que a menudo se expresa en términos de manifestaciones conductuales. Se propone la replicación continua como solución para estas circunstancias, así como para otras situaciones relacionadas con la investigación perimetral (Nosek, 2022). Con la implementación de esta estrategia, la consistencia de los hallazgos se interpreta en el sentido de que proporciona confianza en que los resultados observados en la VD son atribuibles a la manipulación ejercida sobre la VI (Shadish et al., 2002). Otra circunstancia se presenta cuando el investigador carece de un control total sobre el entorno de investigación. En tales casos, el objetivo de la

investigación es observar los cambios en las actitudes que surgen como resultado del desarrollo del programa (González-Garay et al., 2018). Por ejemplo, esto podría implicar la realización de un seminario sobre un tema específico. Sin embargo, el nivel de control que el investigador tiene sobre la situación experimental, como el seminario, puede verse influenciado por varias variables incontrolables. La principal ventaja de la investigación cuasi-experimental es su relativa facilidad y rentabilidad en comparación con los estudios experimentales. Además, es la única opción viable cuando la asignación aleatoria no es factible o cuando es necesario realizar el estudio en condiciones naturales. Sin embargo, estos estudios son muy propensos a sesgos, en particular a la selección y la confusión. Además, el uso de grupos pre-existentes puede comprometer la validez externa y la aplicabilidad de los hallazgos (Shadish et al., 2002).

Por último, en contraste con los estudios cuasi-experimentales, los estudios experimentales ofrecen un mayor control en su diseño, ya que minimizan el sesgo mediante la asignación aleatoria de grupos. También son replicables y comparables a otros estudios, aunque a un costo mayor. Sin embargo, conllevan limitaciones éticas y la responsabilidad de manipular la exposición. La generalización puede ser difícil debido al sesgo de selección o a la naturaleza rígida de la intervención. Este tipo de estudio se emplea con mayor frecuencia en la evaluación de tratamientos para enfermedades específicas entre los pacientes (Campbell & Stanley, 1979). La validez fundamental de este estudio radica en el proceso de aleatorización, que garantiza la comparabilidad de los grupos entre las variables más relevantes relacionadas con el problema de investigación.

3.3. Metodologías empíricas cualitativas

El nivel máximo de complejidad de cualquier teoría reside en la simplicidad de expresar un concepto. A la luz de esto, la investigación cualitativa puede definirse como la investigación de las personas en función de sus acciones verbales y conductuales dentro del contexto social y cultural. Por lo tanto, el objetivo de la investigación cualitativa es comprender las experiencias de las personas desde su propia perspectiva (Taylor & Bogdan, 1987). Según esto, los estudios cualitativos se caracterizan por su capacidad para los investigadores de centrarse en los sujetos, adoptando una perspectiva intragrupo o un enfoque global para abordar el fenómeno en estudio. Según Taylor y Bogdan, los investigadores cualitativos se esfuerzan por comprender las declaraciones de las personas. El acto de llevar a cabo una investigación cualitativa se percibe como algo bastante

sencillo. Existe una idea errónea muy extendida, carente de cualquier base, de que la investigación cualitativa, que se centra en entender lo que dice la gente, es simple, fácil, conveniente y trivial. Existe una serie de preguntas que ayudan a definir cualquier fenómeno dado y a aclarar el enfoque de investigación que se utilizará (Polgar & Thomas, 2021). Estas preguntas buscan localizar el fenómeno, definirlo, aclarar las dimensiones y sus variaciones, la importancia del mismo, su existencia y significado, así como su origen.

Dada la naturaleza extensa de los análisis que se desarrollan, los estudios de investigación cualitativa pueden ser clasificados según sean estudios interpretativos o estudios descriptivos. Por otro lado, los estudios interpretativos buscan una teoría con respaldo mediante un análisis inductivo. Mientras que los estudios descriptivos comprenden diversos diseños que pueden ser documentales, de investigación-acción, biográficos o narrativos, etnográficos, o fenomenológicos. Los diseños etnográficos profundizan en los estudios exhaustivos de un grupo humano específico, que abarcan los aspectos materiales e inmateriales de la cultura (Swanson & Holton, 2005). Esto incluye el examen de diversas facetas, como la economía, la familia y el parentesco, la alimentación, las relaciones sociales, la política, el simbolismo y el idioma. El objetivo principal de los diseños etnográficos es analizar grupos humanos diferenciados cultural o étnicamente. Los métodos de recopilación de datos para estos diseños consisten principalmente en la observación participante y las entrevistas en profundidad. El objetivo es dilucidar y analizar las nociones, conocimientos, convicciones, significados, y prácticas de grupos en distintas culturas. Los diseños etnográficos pueden abarcar aspectos amplios, como la historia, los factores socioeconómicos, la educación, la política, la geografía, y los subsistemas culturales dentro de un sistema social, que abarcan los rituales, los símbolos, los roles sociales, el parentesco, las migraciones, las redes y otros elementos. Esencialmente, la etnografía puede percibirse como una exploración del entorno cultural completo de un grupo humano cultural o étnicamente distinto.

En lo referente a los estudios o los diseños fenomenológicos, utilizan herramientas metodológicas similares para la recopilación de datos, a saber, la observación participante y las entrevistas en profundidad. Swanson y Holton (2005) enfatizan que los diseños etnográficos exploran específicamente categorías, temas y patrones asociados con las culturas. Los investigadores reflexionan sobre las distintas cualidades que distinguen a un grupo o comunidad en particular de los demás. Consideran aspectos como la estructura

del grupo, las reglas que rigen sus operaciones, las creencias compartidas, los patrones de comportamiento exhibidos, los patrones de interacción, las condiciones de vida, las costumbres, los mitos, los ritos y los procesos centrales dentro del grupo o la comunidad (Hernández et al., 2006).

El diseño biográfico o narrativo tiene como objetivo presentar la perspectiva personal de un informante, abarcando hechos, experiencias, opiniones, evaluaciones y conceptualizaciones de su propia vida. Esta forma de descripción tiene sus raíces en la fenomenología y exige que el investigador posea cuatro habilidades clave: observación, escucha, comparación y escritura. Se selecciona a un informante crucial, cuya vida ejemplifica un aspecto particular de interés para el investigador, para incluirlo en el estudio, que puede involucrar a varios informantes según el diseño. Este enfoque se conoce comúnmente como estudio de caso, y los métodos de recopilación de datos suelen incluir entrevistas en profundidad realizadas a través de varios encuentros con el informante. Según el diseño biográfico específico, también se pueden incorporar documentos adicionales, como grabaciones de vídeo, imágenes, cartas privadas y diarios personales, que proporcionen información detallada sobre eventos y experiencias importantes en la vida del informante.

Por otro lado, los diseños de investigación-acción participativa se caracterizan por la participación activa y la postura no neutral del investigador en colaboración con los grupos objeto de estudio, con el objetivo de lograr cambios a través del conocimiento y la acción. Este enfoque hace hincapié en los resultados prácticos al empoderar a los participantes para que aborden los problemas utilizando sus propios recursos, incluidos el conocimiento, la reflexión, la intervención, la acción y la resolución (Creswell & Creswell, 2018; Sampieri & Mendoza, 2018). El diseño reconoce explícitamente la falta de neutralidad tanto del investigador como del proceso de investigación, alineándose con los grupos marginados y posicionando al investigador como un catalizador de la transformación social. Esta investigación se centra en los grupos humanos que se enfrentan a desafíos que requieren la intervención de los investigadores para resolverlos, y los métodos de recopilación de datos incluyen principalmente la observación participante, las entrevistas en profundidad, las entrevistas grupales, los talleres, la revisión de documentos y el trabajo de campo. El objetivo general de la investigación-acción participativa es facilitar cambios concretos en las comunidades mediante el aprovechamiento de sus recursos y su participación y, en última instancia, fomentar la

generación de conocimientos empoderados basados en la sabiduría local. Este proceso empodera a las comunidades, inicia o fortalece estrategias de cambio y tiene como objetivo vincular los esfuerzos locales con iniciativas similares a nivel mundial para impulsar una transformación social más amplia (Marti, 2017).

Por último, el diseño documental abarca la investigación que implica la consulta de diversas fuentes de información registrada. Antes de describir sus modalidades, conviene distinguir dos usos fundamentalmente distintos del documento en investigación, pues esta distinción determina el enfoque metodológico adoptado. En primer lugar, el documento como fuente de información: el investigador lo utiliza para acceder a datos sobre una realidad externa al propio documento. Lo que se estudia no es el documento en sí, sino el fenómeno al que hace referencia. Es el caso de la revisión bibliográfica, el meta-análisis o el uso de bases de datos estadísticas, el documento es un instrumento de acceso al conocimiento acumulado, no el objeto de análisis. En segundo lugar, el documento como objeto de análisis, el investigador estudia el propio documento, su contenido, su estructura, su contexto de producción y la intención de quien lo escribe. Lo que interesa es el documento como producto cultural o comunicativo y el sujeto que lo genera. Es el caso del análisis de contenido, el análisis del discurso, el estudio de textos históricos o la investigación biográfica basada en documentos personales.

Dentro del primer uso, existe un tipo específico conocido como investigación secundaria, que abarca distintas investigaciones bibliográficas y de revisión narrativa, de meta-análisis, meta-síntesis, etc. Independientemente de si la investigación es cuantitativa o cualitativa, el enfoque metodológico de los diseños documentales sigue siendo coherente (Creswell & Creswell, 2018; Hernández-Sampieri & Mendoza, 2018). Sin embargo, los diseños específicos varían según el tema y los objetivos elegidos. En ambos casos, es esencial que la recopilación de datos sea un proceso sistemático con definiciones y especificaciones claras, y que las fuentes se citen correctamente indicando cómo fueron accedidas. Entre los ejemplos de posibles fuentes se incluyen archivos, hemerotecas y bases de datos, y las estrategias de búsqueda y selección de documentos deben especificarse claramente. La investigación documental se basa en la información recopilada de los documentos, entendidos como cualquier material permanente utilizado como fuente o referencia para proporcionar información o explicar una realidad o un evento, sin alterar su naturaleza o significado (Sánchez & Valdés, 2008).

3.4. Metodologías mixtas

Los métodos mixtos corresponden a una colección de metodologías de investigación de orden empírico sistémico con recopilación de datos cualitativos y cuantitativos. El propósito es integrar y discutir estos datos para sacar conclusiones y obtener una comprensión más profunda del fenómeno bajo investigación. Este enfoque hace hincapié en la combinación de metodologías cualitativas y cuantitativas en un solo sistema, con el objetivo de proporcionar una descripción más completa del fenómeno. Hay que reconocer que estos métodos pueden combinarse conservando sus estructuras y procedimientos originales, o adaptarse, alterarse o sintetizarse para abordar los objetivos de la investigación de manera efectiva. En esencia, la metodología mixta incluye datos numéricos, simbólicos, visuales, verbales y textuales, para comprender los problemas científicos (Levitt et al., 2018).

La integración de metodologías cuantitativas y cualitativas en la investigación mixta se considera un procedimiento continuo, en el que se ponen diferentes niveles de énfasis en cada enfoque. La utilización de métodos mixtos solo se considera apropiada cuando ofrece beneficios adicionales a la investigación en comparación con el empleo de una metodología singular (Creswell & Creswell, 2018; Sampieri & Mendoza, 2018). Esto se debe a que, por lo general, requiere mayores recursos económicos, la participación de más personas y más tiempo y conocimientos. Además, las intrincadas características de muchos fenómenos de investigación o cuestiones abordadas en diferentes campos también sirven como fuerza impulsora detrás de la adopción de métodos mixtos. Estos fenómenos consisten en realidades tanto objetivas como subjetivas.

El enfoque mixto ofrece varias ventajas para los investigadores que les permite una mayor comprensión del fenómeno, proporcionando una comprensión más holística y completa. Permite la producción de datos más ricos y diversos al considerar múltiples fuentes, tipos de datos, contextos, entornos y enfoques analíticos. Además, mejora la creatividad teórica mediante procedimientos rigurosos de evaluación crítica. El uso de métodos mixtos también facilita la generalización de los resultados al tener una mayor exploración del fenómeno. Además, enriquece la representación de las muestras, lo que conduce a una mayor validez de los resultados. Esto garantiza su adecuación, utilidad y mejora continua. También garantiza la integridad del tratamiento o la intervención en estudio, lo que mejora su fiabilidad. Por último, la optimización de los significados facilitada por los

métodos mixtos permite una perspectiva más amplia para sus hallazgos (Creswell & Creswell, 2018; Sampieri & Mendoza, 2018).

3.5. Metodologías de Síntesis

Las metodologías de síntesis, revisión narrativa, revisión sistemática, meta-síntesis, meta-análisis y bibliometría, se diferencian de las metodologías de generación de evidencia empírica primaria en que su objetivo es integrar, cuantificar y generalizar los hallazgos acumulados en un campo de estudio. Su uso presupone la existencia de investigación empírica previa producida mediante las metodologías cuantitativas, cualitativas o mixtas descritas en los apartados anteriores, por lo que su papel es auxiliar y de síntesis, más que de aportación empírica genuinamente original. No obstante, su valor para el avance de la ciencia psicológica es capital: las revisiones sistemáticas y los meta-análisis ocupan la posición más elevada en la jerarquía de la evidencia científica (Sánchez-Meca, 2010). Las ciencias del comportamiento y más específicamente la Psicología buscan la evidencia empírica para obtener la información más confiable y respaldada científicamente para su aplicación en la práctica clínica (Gómez- Luna et al., 2022). La comprensión e interpretación de la evidencia científica requieren comprender los diferentes niveles de evidencia disponibles, y las revisiones sistemáticas y/o los meta-análisis de las intervenciones psicológicas ocupan la posición más alta en la jerarquía de evidencia. Sin embargo, existen otras metodologías de revisión utilizadas en Psicología que merecen la pena mencionar por sus objetivos relacionados con la síntesis de información relativa a la documentación sobre un mismo tópico (Rother, 2007).

En esta línea, la revisión bibliográfica con un procedimiento estructurado, conocida como revisión narrativa, tiene como objetivo localizar y recuperar información relevante para responder preguntas relacionadas con diversas prácticas, como la clínica, la enseñanza, la investigación o la gestión (Rother, 2007). Esto implica buscar en las bases de datos utilizando una estrategia específica y presentar la información de manera sintetizada. Por otro lado, la meta-síntesis es una forma de evidencia científica que implica un ejercicio de interpretación complejo para mejorar la relevancia y la aplicabilidad de los estudios cualitativos. El objetivo de la meta-síntesis es proporcionar una comprensión más amplia de los fenómenos mediante la comparación, interpretación e integración constantes de los hallazgos. Si bien las técnicas específicas pueden variar, todas tienen como objetivo lograr una interpretación integral. La meta-síntesis sirve como un medio para ir más allá

del conocimiento específico generado a partir de los estudios cualitativos primarios y avanzar hacia una comprensión más completa de los fenómenos. Desempeña un papel crucial en la documentación de evidencia útil para las disciplinas que pueden aplicarla en la práctica (Carrillo et al., 2008).

Sin embargo, dentro de las metodologías de revisión, ocupan un papel imprescindible las revisiones sistemáticas y los meta-análisis. Por su lado, la revisión sistemática es un componente de las metodologías de revisión que implican examinar la evidencia disponible sobre una intervención en particular para abordar preguntas específicas. Sigue una metodología clara y rigurosa y ha evolucionado hasta convertirse en un diseño de investigación distinto, en el que la atención se centra en los trabajos originales que se revisan en lugar de en los pacientes o las unidades administrativas. Las revisiones sistemáticas son resúmenes bien definidos y estructurados de la información disponible que tienen como objetivo abordar cuestiones clínicas específicas. Como abarcan varios artículos y fuentes de información, se consideran el nivel más alto de evidencia. Una de las características que definen a las revisiones sistemáticas es la capacidad de sintetizar la evidencia existente de una manera comprensible. Como se mencionó anteriormente, las revisiones sistemáticas pueden proporcionar información relativa a dudas relacionadas con tratamientos o intervenciones. La diferencia radica principalmente en los estudios primarios que se incluyen y evalúan para cada tipo de pregunta.

Cuando una revisión sistemática incluye una medida cuantitativa del efecto que se está investigando en cada estudio y aplica técnicas estadísticas para combinar estos efectos, se denomina meta-análisis (Sánchez-Meca, 2010). Por lo tanto, el meta-análisis puede definirse como el agrupamiento de una colección de resultados en un solo estudio con un análisis estadístico unificado e integrador. Un meta-análisis es un tipo de revisión sistemática que analiza estadísticamente los tamaños de los efectos de los estudios, proporciona una estimación conjunta del efecto promedio y evalúa la consistencia u homogeneidad de estos tamaños de efecto (Kelley & Preacher, 2012). En los casos en que los tamaños del efecto muestran heterogeneidad, se realizan análisis adicionales para examinar la influencia de las características del estudio (variables moderadoras) que pueden contribuir a esta variabilidad (Botella & Sánchez-Meca, 2015).

Por último, la utilización de métodos bibliométricos en las revisiones ha ganado popularidad por varias razones, la inclusión de herramientas de software, enfoques

interdisciplinarios y una mayor capacidad para manejar grandes cantidades de datos (Donthu et al., 2021). Los enfoques bibliométricos, que son cuantitativamente objetivos, mitigan el sesgo de selección de muestras en las revisiones sistemáticas. Además, los métodos bibliométricos son muy adecuados para discernir el rendimiento de las revistas, los patrones de coautoría, las tendencias de citación y las tendencias fundamentales de la investigación en campos específicos (Kent et al., 2020). La metodología bibliométrica engloba la aplicación de técnicas cuantitativas (es decir, análisis bibliométricos, por ejemplo, análisis de citas) sobre datos bibliométricos (por ejemplo, número de publicaciones y citas). La proliferación de la bibliometría es bastante reciente, como se observa a través de su crecimiento en el campo de la Psicología. Concretamente, las publicaciones que utilizan la bibliometría han crecido a lo largo de los años, con una media de 1.021 publicaciones en la última década, lo que puede atribuirse al crecimiento de la propia investigación científica. Sin embargo, los grandes conjuntos de datos bibliográficos han hecho que los métodos clásicos de revisión sean engorrosos y poco prácticos (Ramos-Rodríguez & Ruíz-Navarro, 2004). Cabe destacar que las grandes bases de datos permiten que la generación de datos bibliométricos sea relativamente fácil, y los programas informáticos bibliométricos como Gephi, Leximancer y VOSviewer permiten el análisis de dichos datos de una forma muy pragmática, lo que ha aumentado el interés académico por el análisis bibliométrico en los últimos tiempos.

En este punto, es importante comparar el análisis bibliométrico con otras alternativas de revisión utilizadas con frecuencia, como el meta-análisis y las revisiones sistemáticas de la literatura. A diferencia de las revisiones sistemáticas de la literatura, que tienden a basarse en técnicas cualitativas, que podrían verse empañadas por el sesgo de interpretación de estudiosos con diferentes formaciones académicas (Maccoun, 1998), el análisis bibliométrico y el meta-análisis se basan en técnicas cuantitativas y, por tanto, pueden evitar o mitigar ese sesgo. Dado que tanto el meta-análisis como el análisis bibliométrico son de naturaleza cuantitativa, la distinción entre ambos métodos puede resultar confusa para algunos estudiosos (Broadus, 1987). Para arrojar luz sobre esta distinción, se deben tener en cuenta que sus métodos cuantitativos son relativamente diferentes en términos de uso, aunque ambos pueden manejar grandes cantidades de literatura. En concreto, el meta-análisis se centra en resumir la evidencia empírica mediante el análisis de la dirección y la fuerza de los tamaños del efecto y las relaciones entre las variables (Kelley & Preacher, 2012). Por el contrario, el análisis bibliométrico

resume la información de un campo a través del análisis de las conexiones sociales y distintas integraciones de la investigación (Donthu et al., 2021).

En resumen

- En Psicología coexisten cinco grandes enfoques metodológicos: bibliográfico/documental (herramienta de apoyo), empírico cuantitativo, empírico cualitativo, mixto y de síntesis.
- La clasificación más clásica distingue entre metodologías manipulativas (experimental) y no manipulativas (observacional, correlacional). La propuesta de este manual amplía esa clasificación para incluir los enfoques cualitativos y de síntesis, cuya presencia en la investigación psicológica contemporánea es creciente.
- Las metodologías de revisión sistemática y meta-análisis son herramientas de síntesis, no de generación de evidencia primaria: su función es integrar y cuantificar los hallazgos acumulados en un campo.
- La elección del enfoque metodológico suele depender de la pregunta de investigación: no existe un método mejor en abstracto, sino el más adecuado para responder a cada pregunta concreta.

4. El control de las variables extrañas

Cuando un investigador diseña un experimento, su objetivo fundamental es poder afirmar con confianza que los cambios observados en la variable dependiente (VD) son consecuencia de la manipulación de la variable independiente (VI) y no de ningún otro factor ajeno al diseño. Sin embargo, en la práctica, cualquier situación de investigación está atravesada por una multiplicidad de variables que no forman parte del objetivo del estudio pero que pueden influir sobre los resultados. Estas variables, denominadas variables extrañas (VE), representan la principal amenaza a la validez interna de un experimento, y su control constituye uno de los pilares metodológicos de la investigación experimental en Psicología (Campbell & Stanley, 1979; Shadish et al., 2002). La regulación de estas variables no explicativas es, en consecuencia, de suma importancia. En ausencia de control sobre estas variables, no se pueden determinar los factores causales que subyacen a los resultados observados. En la investigación psicológica (aunque no se limita exclusivamente a este campo), cada nivel discreto de la variable independiente engendra un grupo o afección experimental distinto. Por ejemplo, si el objetivo es evaluar la eficacia de un tratamiento novedoso, la variable independiente se clasificaría como tipo de tratamiento, y sus niveles podrían abarcar dos categorías: terapia convencional y terapia nueva. El escenario óptimo para evaluar si la nueva terapia produce un efecto diferente en comparación con el enfoque convencional implicaría que ambos grupos fueran homogéneos en todos los aspectos (con la excepción, por supuesto, del tipo de tratamiento). Si se cumpliera esta condición, cualquier discrepancia observada en los resultados podría atribuirse únicamente a la varianza entre los dos grupos, es decir, al tipo de terapia administrada. Por el contrario, si se determina que los participantes de un grupo muestran una mayor motivación para el cambio (o una menor), son más jóvenes (o mayores), poseen una inteligencia mayor (o inferior), demuestran una mayor perseverancia en las tareas (o una menor) o difieren en cualquier característica que pueda influir de manera plausible en el cambio registrado, surge un problema importante: queda la incertidumbre de si el cambio observado es realmente atribuible a la variable manipulada, a la variable no controlada o a una potencial efecto de interacción derivado de ambos (ya sea sumativo o reflejando los efectos principales) o interactiva [efectos de interacción]). Para resumir: la integridad del experimento se ve comprometida, lo que hace que los datos resultantes sean inválidos (y, por lo tanto, inadecuados para inferir la eficacia de la variable independiente sobre la variable dependiente). ¿Qué estrategias

metodológicas pueden rectificar este problema? La solución está en la aplicación de técnicas de control.

Las técnicas de control son metodologías diseñadas para mitigar el impacto diferencial de variables ajenas en grupos dispares. Es imperativo hacer hincapié en que el objetivo no es necesariamente la erradicación total de la influencia de la variable extraña (lo que, como se demostrará, no siempre es factible o ventajoso), sino más bien garantizar que su impacto se mantenga constante en varios grupos o contextos experimentales. La equivalencia inicial entre los grupos es fundamental, ya que cualquier divergencia que pudiera influir en la variable dependiente impediría determinar si los resultados observados son atribuibles a la variable independiente o a alguna variable alternativa que se correlacione con la composición del grupo. La eliminación es una estrategia según la cual, si es probable que una variable influya en los resultados experimentales y queda fuera del ámbito de interés, lo más lógico sería eliminarla. Por ejemplo, en un experimento de atención auditiva, si es probable que un ruido extraño distorsione los resultados, es prudente realizar el experimento en un entorno libre de ruido durante la fase de recopilación de datos. Sin embargo, es vital reconocer que esto representa simplemente un método para lograr la eliminación. Si mantuviéramos un nivel constante de ruido durante todo el experimento (por ejemplo, mediante la utilización de un generador de ruido blanco), también neutralizaríamos eficazmente el ruido como variable extraña. La razón detrás de esto es que el ruido no fluctuaría (permanecería estático), por lo que no funcionaría como una variable; su efecto sería idéntico en todas las condiciones experimentales. De este modo, se mitigaría la principal preocupación, a saber, que la variable funciona de manera diferente según las condiciones experimentales. Por lo tanto, cualquier variación posterior entre las condiciones no podría atribuirse a esta variable. Esta técnica es relevante en todas las fases de la recopilación de datos, en las que la introducción de la variabilidad podría influir en los resultados finales, aunque de formas que tal vez no se entiendan del todo. La eliminación suele regir el marco de la recopilación de datos, el momento del experimento, la manera en que se transmiten las instrucciones, los materiales utilizados, etc. Sin embargo, este enfoque parece menos adecuado para las variables temáticas. Por ejemplo, si se considera que el sexo es una variable ajena, el empleo de esta técnica requeriría realizar el experimento exclusivamente con participantes masculinos o femeninos. Si bien esto eliminaría de manera efectiva la posible variable de confusión, al mismo tiempo limitaría gravemente la generalización de

los hallazgos y los haría aplicables únicamente a la cohorte del mismo sexo que participó en el estudio. La misma lógica se aplica a variables como la inteligencia, la personalidad y la ansiedad. En estos casos, es más prudente emplear metodologías alternativas, que se aclararán más adelante. Ver Figura 3, árbol de técnicas de control de VE por fuente de variación.

La **constancia** tiene un enfoque metodológico que abarca la distribución equitativa de los valores asociados a la variable que puede generar confusión, garantizando así que la variable objeto de análisis ejerza una influencia uniforme en todos los grupos de participantes (condiciones experimentales). Por ejemplo, si la intención es evaluar el impacto de ver un vídeo (que presenta tres formatos de argumento distintos) en las actitudes hacia políticas económicas específicas, hay que considerar una posible variable de confusión, como la ideología política (caracterizada por cuatro niveles distintos: afinidad por el PP, el Partido Socialista, IU o Podemos). Suponiendo un total de 60 participantes, se necesitaría una cohorte de 20 personas asignadas a cada una de las tres condiciones de vídeo. La aplicación de la técnica de constancia implicaría la asignación equitativa de los niveles de la variable de confusión en cada condición, específicamente, 5 personas alineadas con el PP, 5 personas afiliadas al PSOE, 5 personas asociadas a IU y 5 personas relacionadas con Podemos en la condición experimental inicial. Posteriormente, se aplicaría un procedimiento idéntico a las dos condiciones experimentales restantes. Este arreglo estratégico garantiza que la influencia potencial de la ideología política se distribuya uniformemente en todas las condiciones, lo que permite atribuir definitivamente cualquier discrepancia observada en la aceptación de las medidas económicas a la influencia del vídeo y no a predisposiciones ideológicas preexistentes. Esta técnica resulta particularmente ventajosa cuando las variables de confusión se refieren a características específicas del sujeto, como se ilustra en el ejemplo anterior. En consecuencia, facilita la mitigación de los efectos de una variable de confusión y, al mismo tiempo, preserva la capacidad de generalizar los hallazgos en todos los niveles de la posible variable de confusión.

La técnica del **balanceo** está diseñada para equilibrar la influencia de la variable de confusión y, al mismo tiempo, preservar la representación proporcional de sus valores dentro de la población de todos los grupos experimentales. El objetivo es mejorar la representatividad de los grupos experimentales en relación con sus poblaciones correspondientes (en ciertos aspectos, esto representa un avance con respecto a la técnica

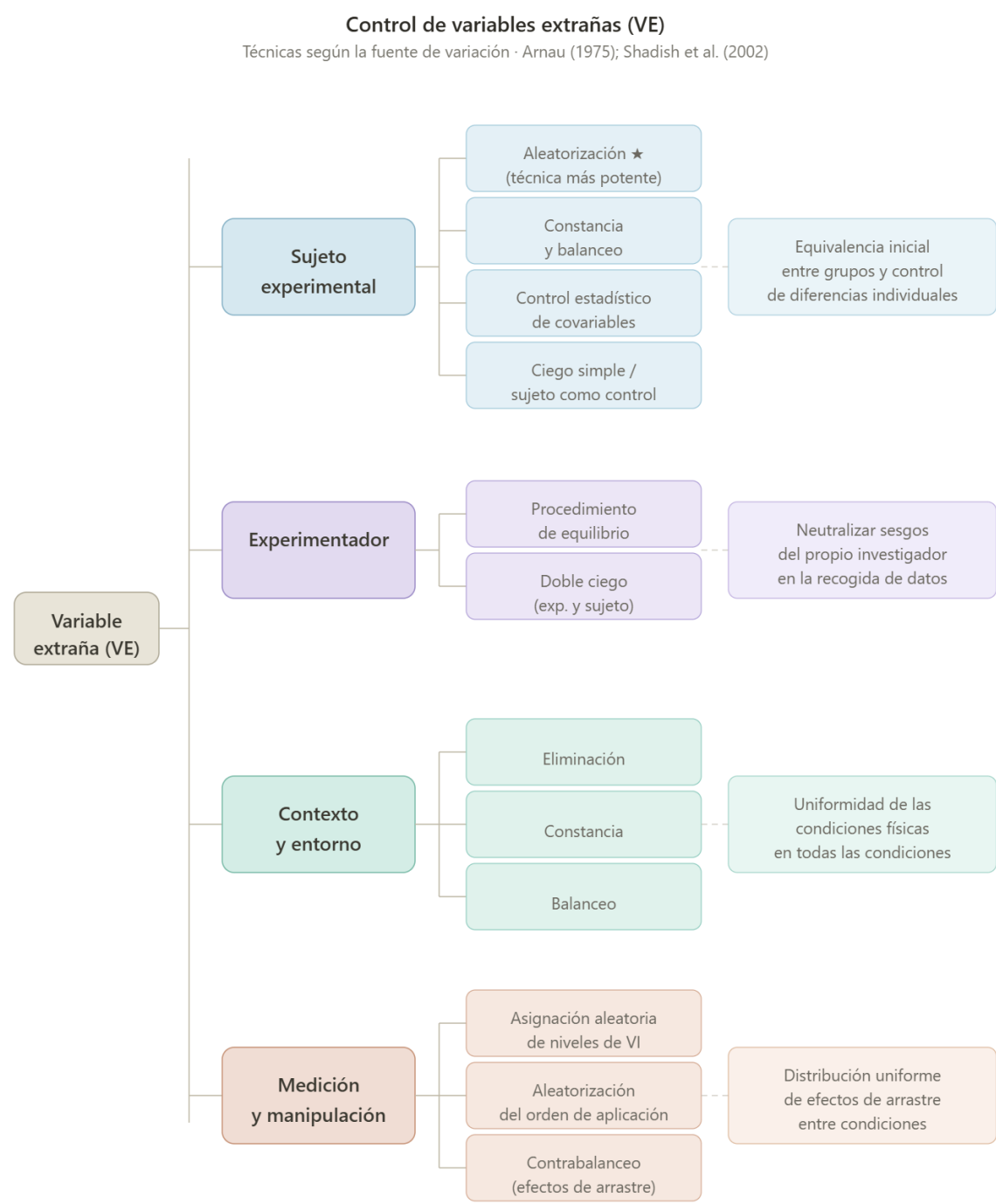
de la constancia). Volviendo al ejemplo antes mencionado, en lugar de asignar un número igual de personas afiliadas a cada una de las cuatro formaciones políticas, se aplicaría una distribución proporcional a la demografía nacional existente. Por ejemplo, dada una distribución del 40% del PP, el 35% de Podemos, el 15% del PSOE y el 10% de IU [en particular, estos porcentajes son hipotéticos y sirven para dilucidar el procedimiento, lo que da como resultado números enteros], primero se determinaría la cantidad de cada grupo (por ejemplo, el 40% de 60 arroja 24) y, posteriormente, se dividiría esa cifra por el número de condiciones (por ejemplo, $24/3 = 8$). Por lo tanto, la distribución revisada de los valores asociados a la variable de confusión estaría compuesta por 8 personas del PP, 7 de Podemos, 3 del PSOE y 2 de IU en cada una de las condiciones experimentales.

Con el enfoque de la **aleatorización** se pretende facilitar la asignación estocástica de variables ajenas en todos los grupos experimentales. Al realizar una investigación relacionada con el aprendizaje, si la variable inteligencia se presenta como una posible variable de confusión, la asignación aleatoria de los sujetos a los grupos debería garantizar una distribución uniforme de esta variable en todos los grupos. Para ello es necesario partir del supuesto de que la aleatoriedad determinará la dispersión de las personas con altos niveles de inteligencia entre los distintos grupos, así como de las personas con niveles de inteligencia más bajos. La probabilidad de que los diez individuos más inteligentes sean asignados a un solo grupo experimental (o, por el contrario, los diez menos inteligentes) por pura casualidad es extremadamente baja. Esta metodología prevalece en la investigación empírica; sin embargo, es imperativo reconocer que su eficacia no siempre se ajusta a las expectativas, ya que depende del tamaño de la muestra. A medida que aumenta el número de participantes y disminuye el número de grupos, aumenta la capacidad de distribución aleatoria para lograr su objetivo de distribuir las posibles variables de confusión. La asignación de 20 participantes en 5 grupos difiere fundamentalmente de la asignación de 100 participantes en 2 grupos. En el primer escenario, la probabilidad de que dos de cada cuatro sujetos del mismo grupo posean valores extremos en la variable de confusión es considerablemente alta, lo que se traduce en una composición desigual de los grupos. Por el contrario, en el segundo escenario, incluso si se produjera tal situación (2 valores extremos entre los 50 sujetos del mismo grupo), las condiciones relativas a la variable de confusión (inteligencia) seguirían siendo prácticamente equivalentes.

Por su otra parte, el **contrabalanceo** es una técnica que sirve para mitigar la influencia

de los efectos del error progresivo (efectos de arrastre), que prevalecen en los diseños experimentales en los que los participantes están expuestos a múltiples condiciones experimentales (estas afecciones abarcan los efectos de la fatiga, los efectos del aprendizaje o los efectos residuales). El objetivo es garantizar que el nivel de error acumulado (fatiga, aprendizaje, etc.) se distribuya uniformemente en cada condición experimental. Dada la naturaleza extensiva de este tema y su próxima exploración en cursos posteriores, no se proporcionará más información sobre esta técnica.

Figura 3. Árbol de técnicas de control de VE por fuente (sujeto, experimentador, contexto, medición), con cuadro resumen de objetivos.



★ La aleatorización es la técnica más potente y de mayor uso (Arnau, 1975; Campbell y Stanley, 1979)
VI = variable independiente · VD = variable dependiente · VE = variable extraña

5. Diseños de investigación

Diseño de investigación: plan estructurado que especifica cómo se seleccionan los participantes, qué variables se observan o miden, en qué secuencia y condiciones se realizan las observaciones, y cómo se analizarán los datos. En los diseños con estrategia manipulativa, el plan incluye además cómo se asignan los participantes a las condiciones y qué variables se manipulan y controlan. El diseño determina el tipo de inferencia que es posible extraer de los datos: los diseños observacionales y de encuesta permiten describir y explorar relaciones entre variables; solo los diseños experimentales permiten establecer relaciones causales.

Diseño experimental: diseño en el que el investigador (a) manipula al menos una VI, (b) asigna los participantes aleatoriamente a las condiciones, y (c) controla las VE mediante procedimientos sistemáticos. Es el único diseño que permite inferir relaciones causales entre variables (Campbell & Stanley, 1979).

El presente apartado se organiza en torno a los diseños de investigación más relevantes para la metodología experimental en Psicología. El núcleo del capítulo lo constituyen los diseños experimentales, aquellos en los que el investigador manipula la VI y asigna aleatoriamente los participantes a las condiciones, por ser los únicos que permiten inferencia causal directa. Los diseños cuasi-experimentales se presentan a continuación como una aproximación al ideal experimental cuando las condiciones de laboratorio no son viables (por razones éticas o pragmáticas), y se exponen con el nivel de detalle suficiente para que el estudiante reconozca sus posibilidades y sus limitaciones. Los diseños longitudinales y otros diseños avanzados (factoriales incompletos, series temporales, Solomon) se mencionan brevemente como extensiones para contextos específicos y se remiten a bibliografía especializada para su estudio en profundidad (Kirk, 2013; Maxwell et al., 2018).

El diseño de investigación es el plan que establece cómo vamos a recoger los datos para responder a nuestra pregunta. La elección del diseño determina directamente qué tipo de conclusiones podemos extraer. Este apartado se organiza en torno a una jerarquía clara: los diseños experimentales constituyen el núcleo, por ser los únicos que permiten establecer relaciones causales. Los diseños cuasi-experimentales se presentan como alternativa cuando las condiciones del experimento puro no son viables. Los diseños

longitudinales y otros diseños avanzados se mencionan como extensiones para contextos específicos y se remiten a bibliografía especializada (Shadish et al., 2002; Kirk, 2013).

Ejemplo: Queremos saber si el café mejora la concentración. VI: tipo de bebida (con cafeína / sin cafeína / agua). VD: puntuación en una tarea de atención sostenida de 20 minutos. Procedimiento: asignamos aleatoriamente a los participantes a una de las tres condiciones; les administramos la bebida de forma ciega (sin que sepan cuál es); medimos la concentración con el mismo instrumento en todos los casos. Este diseño cumple los tres requisitos del experimento: (1) manipulación de la VI, (2) asignación aleatoria a condiciones, y (3) control de VE (mismo contexto, mismo instrumento, mismo evaluador). Con este diseño podemos afirmar causalmente que el café afecta, o no, a la concentración.

En resumen

- Una variable extraña (VE) es cualquier factor distinto de la VI que puede influir sobre la VD y que, si no se controla, amenaza la validez interna del experimento.
- El objetivo del control no es eliminar la VE (a menudo imposible), sino garantizar que su impacto sea equivalente en todos los grupos experimentales.
- Las técnicas de control se clasifican según su fuente: del sujeto experimental (aleatorización*, constancia, balanceo, control estadístico, ciego simple), del experimentador (equilibrio, doble ciego), del contexto (eliminación, constancia ambiental) y de la medición (aleatorización del orden, contrabalanceo).
- La aleatorización es la técnica más potente: al asignar aleatoriamente los participantes a las condiciones, las VE tienden a distribuirse de forma equilibrada entre los grupos, neutralizando su efecto diferencial.

En la categorización de varios diseños de investigación, se postula una alineación con una estrategia específica, que depende del objetivo perseguido. Entre las formas fundamentales del diseño de investigación, cada diseño individual posee características distintivas que lo diferencian de los demás. Sin embargo, debe reconocerse que, en la práctica, los diseños raramente se aplican en su forma más pura, y con frecuencia ciertas características se adaptan o combinan para ajustarse mejor a las condiciones y objetivos de una situación de investigación determinada (Ato et al., 2013). Posteriormente,

expondremos los diversos diseños de investigación, basados en la presencia de la manipulación de la variable independiente, en una estrategia de interrelación entre las variables o en una estrategia que dilucide el estado de las variables. Principalmente, describiremos los diseños de investigación en los que la estrategia de manipular la variable arroja una metodología experimental o una metodología cuasi-experimental (Balluerka & Vergara, 2002). Por otro lado, describiremos los diseños de investigación que emplean estrategias asociativas, en las que se puedan encontrar diseños comparativos, diseños predictivos o diseños explicativos entre variables. Por último, abordaremos los diseños de investigación que se relacionan con las estrategias descriptivas de las variables, que abarcan los diseños observacionales o los diseños de encuestas.

5.1. Diseños con estrategia manipulativa

5.1.1. Diseños experimentales

Existen diferentes perspectivas en la conceptualización de los diseños experimentales. El punto de vista global sugiere que el diseño experimental abarca todos los aspectos de la metodología, incluida la especificación de la población, la selección de la muestra, la definición de variables (VI, VD, VE), la formulación de hipótesis, la asignación de sujetos a diferentes condiciones y el análisis estadístico (Ato & Vallejo, 2007). Por otro lado, el enfoque reduccionista, considera que el diseño experimental es un componente crucial de la metodología, que implica la selección del sujeto, la asignación de condiciones y el control meticuloso de las VE para establecer vínculos causales entre la VD y el VI.

Antes de profundizar en los tipos de diseños experimentales, es importante tener en cuenta que la naturaleza experimental de un diseño viene determinada por el cumplimiento de condiciones específicas. La falta de control de las VE o la falta de aleatorización en la asignación de las asignaturas clasifican un diseño como cuasi-experimental. Además, si el VI no puede manipularse intencionalmente, el diseño se considera no manipulativo (Martínez, 1995).

Según el recuento del VI, los diseños se clasifican en unifactorial (una VI) o factoriales (dos o más VI). El diseño unifactorial más simple es el diseño de dos grupos aleatorios, en el que una única VI tiene dos niveles. Los diseños multigrupales implican más de dos niveles en la VI. En cuanto al recuento de VD, los diseños se etiquetan como univariados (una VD) o multivariantes (dos o más VD). En términos de comparación de tratamientos,

se identifican los diseños intersujeto, intrasujetos, y mixtos. Los diseños interdisciplinarios comparan los efectos del VI en diferentes sujetos o grupos, asumiendo su equivalencia inicial. Los diseños intrasujetos involucran a un grupo sometido a todos los niveles de VI. Los diseños mixtos combinan elementos intersujeto e intrasujetos, explorando ambos niveles del VI (Pascual et al., 1995)

El criterio de clasificación diferencia entre diseños factoriales completos y diseños factoriales incompletos, basados en las combinaciones de condiciones experimentales. Los diseños factoriales completos incluyen todas las combinaciones posibles de los niveles del VI, mientras que los diseños factoriales incompletos omiten ciertas combinaciones que se consideran menos importantes para eliminar el efecto de algunos VE (Blanca & Rando, 2005). Por otro lado, el parámetro temporal de la VD se refiere al momento del registro, que puede ser en un momento específico o en momentos diferentes a lo largo de un período. Los diseños transversales implican registrar el comportamiento de diferentes grupos de sujetos simultáneamente, mientras que los diseños longitudinales registran múltiples medidas de variables a lo largo del tiempo en el mismo grupo de sujetos (Carlson & Schmidt, 1999). Asimismo, el criterio técnico para la formación de grupos se basa en la técnica de control de VE utilizada para formar grupos experimentales, como los diseños de grupos aleatorios, los diseños de bloques y los diseños dentro de una clasificación en los que el sujeto actúa como control. Por último, el ajuste estadístico implica el uso de la técnica ANCOVA para controlar los posibles VE (Pascual et al., 1995). Los diseños pueden incluir covariables cuando se utiliza ANCOVA para el control de VE, o diseños sin covariables cuando no se aplica esta técnica. Por último, el grado de control de los VE distingue entre diseños experimentales y cuasi-experimentales. Los diseños experimentales implican la asignación aleatoria de sujetos y la asignación aleatoria a niveles VI, mientras que los diseños cuasi-experimentales carecen de aleatorización, lo que reduce el control sobre los VE. El uso de técnicas de aleatorización determina en última instancia si un diseño se considera experimental o cuasi-experimental (Ato & Vallejo, 2015).

5.1.2. Diseños cuasi-experimentales

Los diseños cuasi-experimentales sirven como alternativa a los diseños experimentales reales cuando no es posible el control total sobre las VE o no es posible la asignación aleatoria de sujetos a grupos experimentales o niveles de VI. Estos diseños se pueden

clasificar atendiendo a dos criterios independientes: según la forma en que se distribuyen los participantes entre las condiciones, distinguiendo entre diseños intersujeto, intrasujetos y mixtos, y según el número de momentos de medición, distinguiendo entre diseños transversales, con una única recogida de datos, y diseños longitudinales, con mediciones repetidas a lo largo del tiempo (Ato & Vallejo, 2007). Los diseños cuasi-experimentales más utilizados en la investigación psicológica se analizan a continuación, utilizando estos criterios de clasificación (Martínez, 1995).

Dentro de los diseños cuasi-experimentales intersujeto, el diseño fundamental es el diseño de grupos de control no equivalentes, propuesto originalmente por Campbell y Stanley (1963) y desarrollado posteriormente por Cook y Campbell (1979), en el que los grupos se forman de manera no aleatoria, lo que genera incertidumbre sobre su equivalencia u homogeneidad (León & Montero, 2004). Esto plantea la preocupación de que las diferencias observadas entre los grupos después del tratamiento puedan deberse al tratamiento en sí o a diferencias grupales preexistentes. Para abordar este problema, las pruebas previas y el análisis de covarianzas pueden emplearse como técnicas de control estadístico. Este enfoque puede ampliarse a escenarios que involucren más de dos niveles de administración de la condición experimental, con o sin un grupo de control (Kirk, 2013).

Entre los diseños cuasi-experimentales intrasujeto, se describen cuatro tipos distintos. Estos incluyen los diseños de un solo grupo antes y después de la prueba (*one-group pretest-posttest design*), introducido por Campbell y Stanley (1963) y considerado uno de los diseños cuasi-experimentales más básicos, en los que la VD se mide antes y después de la aplicación del tratamiento para evaluar los efectos del tratamiento. El intervalo de tiempo entre la fase previa y la posterior a la prueba es fundamental, ya que los intervalos más largos aumentan la probabilidad de que factores externos influyan en las diferencias observadas. Varias amenazas a la validez interna de este diseño, como la historia, la maduración, la regresión estadística y la instrumentación, sistematizadas por Shadish et al. (2002), pueden mitigarse mediante modificaciones, como la utilización de múltiples medidas previas a la prueba y diseños no equivalentes de VD antes y después de la prueba (Ato & Vallejo, 2015).

Los diseños de discontinuidad en la regresión (*regression discontinuity design*), desarrollados por Thistlethwaite y Campbell (1960) y posteriormente sistematizados por

Cook y Campbell (1979), incluyen grupos experimentales y de control formados en función de un límite de puntuación antes de la prueba, y el tratamiento aplicado en consecuencia. A continuación, se comparan las mediciones realizadas después de la prueba, con diferentes líneas de regresión que indican la eficacia del tratamiento. Los diseños de series temporales interrumpidas (*interrupted time series design*), introducidos por Campbell y Stanley (1963) y ampliados por Shadish, Cook y Campbell (2002), implican la toma de medidas ordenadas a lo largo del tiempo, lo que permite detectar los efectos de la intervención al observar los cambios en la VD antes y después del tratamiento (Ato & Vallejo, 2015). También se examinan los patrones de cambio posteriores a la intervención en función del momento y la forma de inicio, distinguiendo entre cambios bruscos y graduales. En un esfuerzo por examinar más a fondo estos tipos de diseños de una manera más detallada, Vallejo (1995) introdujo una clasificación que distinguía cinco tipos de diseños de series temporales. Estos incluyen: (a) Diseños de series temporales ininterrumpidas, que reflejan el diseño antes mencionado; (b) Diseños de series temporales con un grupo de control no equivalente, similar al tipo anterior pero con la adición de un grupo de control para mejorar el poder de inferencia con respecto a los efectos del tratamiento; (c) Diseños de series temporales interrumpidos con variables dependientes no equivalentes, en los que las evaluaciones previas y posteriores se realizan en dos variables dependientes diferentes para observar los cambios causados por el tratamiento en una variable y mantener constante la otra; (d) Los diseños de series temporales interrumpidas con réplicas múltiples, que implican mediciones alternas previas y posteriores de la variable dependiente cuando se aplica y se retira el tratamiento; (e) Diseños de series temporales interrumpidas con réplicas intercambiadas, en los que dos grupos experimentales se someten al tratamiento alternativamente y se registran antes y después de cada aplicación para mejorar la validez interna y externa (Cook et al., 1990).

Los diseños de un solo caso se caracterizan por una estructura de investigación que consiste en observaciones repetidas en condiciones o fases uniformes. Inicialmente, la fase inicial (línea basal) o sin tratamiento (fase A) implica registrar las respuestas de los sujetos sin tratamiento hasta que surja un patrón de comportamiento estable. Posteriormente, se implementa el tratamiento (fase B) para evaluar cualquier cambio de comportamiento (Maxwell et al., 2018). Si bien algunos autores clasifican los diseños de un solo caso como cuasiexperimentales debido a la falta de aleatorización, su objetivo principal sigue siendo similar al de los diseños experimentales: confirmar los cambios

inducidos por el tratamiento bajo un control estricto de variables ajenas (Arnau, 1986; Barlow & Hersen, 1984; Campbell & Stanley, 1963; Kazdin, 1992). Según la reversibilidad de la respuesta, los diseños de un solo caso se clasifican en diseños de reversión y sin reversión. Los primeros muestran que el comportamiento vuelve a los niveles basales después de la eliminación de la intervención y los segundos indican cambios de comportamiento permanentes. En particular, los diseños de un solo caso ofrecen flexibilidad en cuanto a los participantes, las variables y los procedimientos de tratamiento.

Entre los diseños cuasi-experimentales mixtos, destaca el diseño pre-post tratamiento con un grupo de control no equivalente. Este diseño implica el registro previo a la intervención de las variables dependientes tanto en el grupo de control como en el experimental, seguido del registro posterior a la intervención únicamente en el grupo experimental (Pascual et al., 1995). La falta de una selección aleatoria conduce a que los grupos estén intactos y a posibles amenazas para la validez interna, como los efectos de la interacción entre la selección y el historial, la selección y la maduración y la administración (Reid et al., 2010). Para abordar estos desafíos, y mejorar la capacidad de deducción, se han propuesto varias alternativas (Vallejo, 1991). Los diseños previos y posteriores a la prueba con un grupo de control implican múltiples medidas previas a la prueba de la VD antes de la aplicación del tratamiento, para observar las diferencias entre los grupos antes de la intervención, con el objetivo de controlar la regresión estadística y las amenazas de maduración. Sin embargo, este enfoque puede introducir amenazas relacionadas con la administración del tratamiento. Los diseños alternativos utilizan muestras separadas antes y después del tratamiento para eliminar la amenaza de la administración del tratamiento, lo que exige la equivalencia entre el grupo experimental y el de control en ambas mediciones. Por otro lado, los diseños previos y posteriores al tratamiento con un grupo de control pueden utilizar diferentes instrumentos de medición del VD cuando no es posible utilizar instrumentos idénticos la evaluación antes y después del tratamiento (Martínez-Garza, 1988). Asimismo, los diseños de cohortes incluyen grupos de sujetos, con experiencias similares, que se supone que son comparables a la hora de evaluar las diferencias previas y posteriores al tratamiento. Por último, los diseños de cuatro grupos de Solomon combinan diseños posteriores al tratamiento con grupos de control y anteriores y posteriores al tratamiento, creando cuatro grupos para la administración del

tratamiento (Solomon, 1949). Este diseño permite verificar los efectos previos al tratamiento sobre los resultados posteriores al tratamiento.

Nota didáctica: Este tipo de diseño excede el nivel introductorio de este manual. Para un tratamiento detallado se recomienda consultar Shadish et al. (2002) y Kirk (2013).

5.2. Diseños con estrategia no manipulativa

Dentro de los diseños no manipulativos se incluyen los diseños de encuesta y los diseños observacionales. La clasificación de los diseños de encuestas se basa en una dimensión temporal, que distingue entre diseños transversales y longitudinales (Ato et al., 2013). Los diseños transversales recopilan datos de una población al mismo tiempo para describir las diferencias entre grupos con características diferentes (Blanca, 2004), lo que resulta adecuado para estudiar variables relativamente estables en un momento específico. En el ámbito de la salud, estos diseños miden la prevalencia de los trastornos y realizan estudios analíticos sobre los subgrupos (por ejemplo, el género o el origen étnico), con la ventaja de que permiten ahorrar tiempo gracias a una única medición. Los diseños longitudinales implican el registro de variables durante un período prolongado, lo que resulta ideal para que la investigación en Psicología del desarrollo haga un seguimiento de los cambios de comportamiento a lo largo del tiempo mediante múltiples encuestas (Ato et al., 2013; Blanca, 2004). Asimismo, los diseños de cohortes longitudinales secuenciales combinan elementos transversales y longitudinales mediante el estudio de las cohortes longitudinalmente y la adición de nuevas cohortes de forma secuencial (Ato et al., 2013). Los distintos diseños sirven para diferentes propósitos: encuestas de panel para analizar los cambios y encuestas de población para comparar diferentes temas antes y después de las mediciones.

Las encuestas de panel evalúan a los sujetos a lo largo del tiempo para analizar los cambios y las posibles explicaciones, lo que resulta adecuado para estudiar los cambios en el desarrollo y los impactos psicológicos. Por ejemplo, ver si un mismo estudiante aumenta su nivel de estrés en época de exámenes. Sin embargo, las amenazas a la validez interna y externa limitan la generalización de los resultados (Martín, 2011). Los diseños de muestreo poblacional o de series temporales tienen como objetivo superar limitaciones como la mortalidad y reactividad experimentales mediante la repetición de encuestas transversales en diferentes momentos con muestras variables. Los diseños de cohortes se

utilizan habitualmente para estudiar procesos evolutivos dentro de una misma población que comparte una característica común. Emplean una estrategia de encuesta mixta que combina enfoques transversales y longitudinales, que a menudo se consideran estudios prospectivos, pero que también pueden ser retrospectivos. Por su parte, los diseños observacionales distinguen entre estructuras transversales y longitudinales en función del registro temporal, las unidades de observación y los criterios del nivel de respuesta (Martínez, 1995). En relación con el aspecto temporal de los registros, existen dos categorías de registros: los registros únicos en los que las observaciones se realizan en un momento específico y los registros de seguimiento en los que las observaciones se realizan durante un período definido. Es esencial diferenciar entre la recopilación de datos actuales y retrospectivos en cuanto al momento en que se evalúan los registros, como el material de archivo con fines estadísticos o los datos censales. Sin embargo, la coherencia de los criterios es crucial para recopilar, categorizar y controlar las variables que podrían sesgar las observaciones (Anguera et al., 2011; Anguera et al., 2001; Ato et al., 2013).

En cuanto a las unidades de observación, se distingue claramente entre las ideográficas, que se centran en un solo tema, que implican observar una muestra representativa de sujetos (Ato et al., 2013). Fernández y Fernández (1999) han profundizado en esta clasificación, afirmando que la observación dentro del mismo contexto se denomina diseño ideográfico, mientras que la observación en diferentes contextos se denomina diseño nomotético.

Por último, el nivel de respuesta o dimensionalidad clasifica los diseños observacionales: unidimensionales (cuando solo se considera un nivel de respuesta), por ejemplo, imaginemos que queremos estudiar la frecuencia de conductas disruptivas en una clase de primaria. El investigador observa durante 30 minutos una clase, solo registra una única variable el número de veces que los alumnos interrumpen (hablar sin permiso, levantarse, etc.). Y multidimensionales (cuando se tienen en cuenta varios niveles de respuesta), ejemplo el investigador observa durante 30 minutos una clase y registra varios niveles de respuesta simultáneamente: conductas disruptivas (interrumpir, levantarse), conductas académicas (atender, tomar apuntes), estado emocional aparente (aburrimiento, interés), e interacciones sociales (hablar con compañeros, cooperación). La combinación de estos tres criterios da como resultado ocho tipos distintos de diseños observacionales, como el diseño puntual ideográfico unidimensional, donde se observa el comportamiento de un individuo o grupo pequeño en un momento específico, registrando solo un nivel de

respuesta (Ato et al., 2013).

Nota didáctica: Este tipo de diseño excede el nivel introductorio de este manual. Para un tratamiento detallado se recomienda consultar Shadish et al. (2002) y Kirk (2013).

En resumen

- El diseño experimental (manipulación + aleatorización + control de VE) es el único que permite inferir relaciones causales entre variables.
- El diseño cuasiexperimental se diferencia en que el investigador no tiene control total: no hay asignación aleatoria completa o la VI no puede manipularse libremente. Permite inferencias causales más débiles.
- Los diseños sin manipulación (observacionales, de encuesta) permiten describir y explorar relaciones, pero no establecer causalidad.
- La elección del diseño depende de la pregunta de investigación, las posibilidades éticas y prácticas, y el nivel de certeza causal que se pretende alcanzar.

6. Validez

Validez: grado en que una investigación mide lo que pretende medir. En metodología experimental se distinguen cuatro tipos fundamentales (Shadish et al., 2002):

Validez de constructo: ¿miden realmente los instrumentos utilizados los constructos teóricos que se pretende medir?

Validez interna: ¿son los cambios observados en la VD atribuibles a la manipulación de la VI y no a factores extraños?

Validez externa: ¿son generalizables los resultados a otras poblaciones, contextos o momentos temporales?

Validez de conclusión estadística: ¿son correctas las inferencias estadísticas extraídas de los datos?

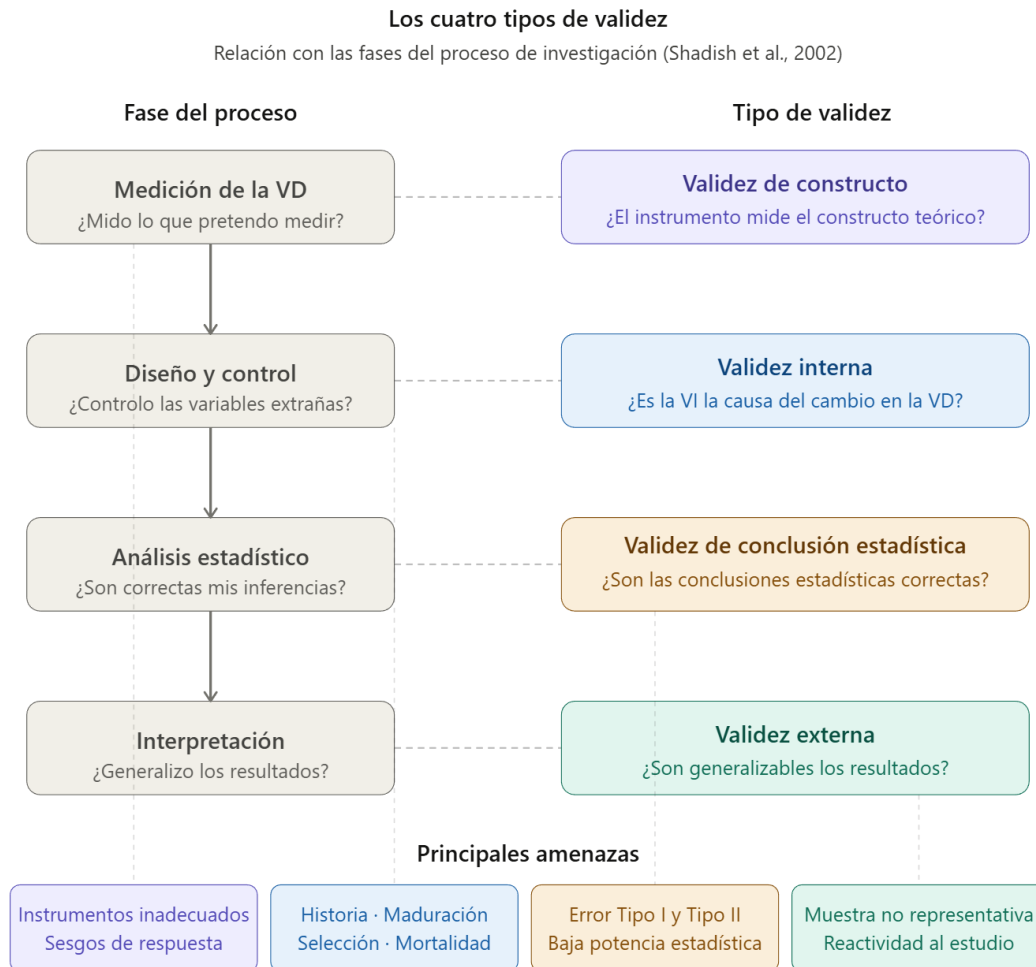
Cualquier forma de investigación está estructurada para obtener información sobre las relaciones funcionales entre las variables que se examinan. El grado en que se alcanza este objetivo depende de la validez del estudio, entendida como la adecuación entre lo que se pretende conocer y lo que realmente se mide, controla e infiere. Aunque la validez se expone en este manual como sección específica por razones de organización didáctica, conviene tener presente que sus cuatro dimensiones son transversales al proceso de investigación y se vinculan cada una de ellas a una fase concreta: la validez de constructo es relevante en el momento de operacionalizar y medir la variable dependiente; la validez interna lo es durante el diseño y el control experimental; la validez de conclusión estadística entra en juego en el análisis de los datos; y la validez externa se evalúa en la fase de interpretación y generalización de resultados (Shadish et al., 2002). El lector puede consultar el Figura 4 para una representación visual de estas correspondencias.

La validez en la investigación se define como el nivel de confianza que se puede depositar en la precisión o exactitud de una investigación (Ato & Rabadán, 1991). Por lo tanto, hablamos de validez en términos de calidad y de las garantías científicas que los resultados encontrados nos aportan (Delgado, 2011). En línea con la perspectiva de Campbell y Stanley (1966), destacan dos categorías fundamentales de validez: la validez interna (sesgos relacionados con el control de la asignación aleatoria) y la externa (sesgos no relacionados con la asignación aleatoria), aunque se incluyen la validez de constructo (precede a la validez externa) y la validez de la conclusión estadística (previa a la validez interna) desarrolladas posteriormente por Cook y Campbell (1979). La evaluación de la validez de la investigación se basa en la calidad de la referencia (una mayor causalidad

implica una mayor validez interna) y en el grado de su generalización (validez externa). Shadish et al. (2002), basándose en el trabajo de académicos anteriores, proponen un marco de validez de la investigación que abarca estos cuatro tipos de validez. Argumentan que la validez de las conclusiones estadísticas y la validez interna están asociadas con la inferencia causal, mientras que la validez del constructo y la validez externa están vinculadas a la generalización de dicha inferencia (Ato & Vallejo, 2007).

A continuación, profundizaremos en cada uno de estos cuatro tipos de validez y sus respectivos desafíos, considerando las amenazas a la validez como aquellos elementos de la investigación en los que existe la probabilidad de que el investigador cometa errores al extraer inferencias causales o emplee indebidamente los constructos y variables en el estudio. Es crucial reconocer la interconexión de los cuatro tipos de validez, ya que una sola amenaza puede afectar a múltiples formas de validez.

Figura 4: Los cuatro tipos de validez articulados con las fases del proceso de investigación.



6.1. Validez de Constructo

Esta forma particular de validez se refiere al grado de acuerdo entre la manipulación de la variable independiente y la medición de la variable dependiente, así como a su conexión con el constructo teórico en estudio. Específicamente, aborda la medida en que los constructos teóricos pueden inferirse a partir de las relaciones causa-efecto entre las variables examinadas en la investigación. Nuestras variables experimentales pueden resultar de la operacionalización de un concepto o constructo, lo que plantea la cuestión de si estas operacionalizaciones reflejan realmente el constructo pretendido. Algunos estudiosos, como Balluerka (1999), diferencian entre dos formas de validez del constructo: la validez del constructo de la causa y la validez del constructo del efecto. La primera se refiere a la medida en que la variable independiente representa el constructo teórico al que se atribuye el efecto del tratamiento sobre la conducta, mientras que la segunda se centra en qué tan bien la variable dependiente representa el atributo teórico

que se está midiendo. La integridad de esta validez se ve comprometida por la falta de claridad o definición del constructo, lo que lleva a una operacionalización o medición inadecuadas, lo que puede mitigarse mediante el desarrollo de un marco teórico sólido. Los participantes en la investigación son humanos por lo que sus reacciones ante los procedimientos experimentales pueden afectar a la validez de constructo y a la validez externa. Técnicas como los estudios de ciego simple, doble ciego y triple ciego ayudan a mitigar este problema al garantizar que tanto los participantes como los experimentadores desconozcan las condiciones experimentales asignadas.

6.2. Validez Interna

La validez interna se refiere a la probabilidad de obtener conclusiones precisas sobre el impacto de la variable independiente en la variable dependiente. Se considera causal, ya que se centra en identificar los factores responsables de la alteración observada en la variable dependiente dentro de un contexto y un período de tiempo específicos. La base fundamental en la que se basa la noción de validez interna es que una deducción causal será fiable en la medida en que el diseño de investigación utilizado sea competente para establecer una conexión temporal entre las variables independientes y dependientes, al tiempo que elimina todas las hipótesis explicativas contrapuestas. Esto asegura que el efecto observado (variable o criterio dependiente) se atribuya únicamente a la variable independiente o predictiva (Ato & Vallejo, 2015). Para lograrlo, además de gestionar las variables externas, es crucial discernir la dirección de la causalidad, ya sea de la variable manipulada (causa) a la variable medida (efecto observado) o viceversa. Esta determinación de la dirección de la causalidad depende de la secuencia cronológica de las variables. Esta diferenciación es relativamente sencilla cuando se emplea el enfoque experimental en las investigaciones, y se complica con los métodos no experimentales (Ato, 1991; Ato & Rabadán, 1991; Ato & Vallejo, 2007). La validez interna está vinculada a la integridad del experimento y se logra controlando la varianza sistemática secundaria, garantizando que los grupos únicamente varíen en el tratamiento y validando la medición de la variable dependiente. Un diseño válido internamente se logra cuando las disparidades observadas en la variable dependiente entre grupos variados pueden atribuirse únicamente a las variaciones inducidas en la variable independiente. Para lograrlo, es imprescindible abordar las siguientes amenazas a la validez interna: el origen temporal incierto de la causa, las influencias históricas, la maduración, las condiciones de prueba, la instrumentación, la regresión estadística, la selección diferencial, el desgaste

experimental y las posibles interacciones de estos factores con una selección variada (Shadish et al., 2002). A continuación, se examinará cada amenaza en detalle.

En los casos de ambigüedad de proximidad temporal con respecto a la causa, la amenaza surge de la falta de claridad a la hora de distinguir la causa del efecto, debido a la ambigüedad de la contigüidad temporal entre las variables examinadas. Ejemplo: Un investigador desea analizar la relación entre el uso de redes sociales y los niveles de ansiedad en adolescentes. Para ello, mide ambas variables en un único momento, encontrando que aquellos con mayor uso de redes sociales presentan mayores niveles de ansiedad. Sin embargo, este diseño presenta ambigüedad de proximidad temporal, ya que no permite determinar si el uso intensivo de redes sociales es la causa del aumento de la ansiedad, si son los adolescentes con mayor ansiedad quienes tienden a utilizar más estas plataformas, o si ambas variables están influidas por un tercer factor no considerado (por ejemplo, problemas previos de ajuste social). En este caso, la falta de información sobre el orden temporal de aparición de las variables impide establecer una relación causal clara. Esta hipótesis suele darse en algunos estudios no experimentales, en particular en las investigaciones realizadas a posteriori, en los que las variables de estudio se refieren a la selección de valores. Resulta difícil determinar la secuencia en la que aparecen las variables (causa o variable independiente) y siguen (efecto o variable dependiente). Para mitigar esta amenaza en las investigaciones no experimentales, se pueden adoptar diseños longitudinales, junto con técnicas analíticas como el análisis de rutas y las ecuaciones estructurales, con el objetivo de establecer la contigüidad temporal entre las variables de investigación (Shadish et al., 2002).

La historia se puede definir como los acontecimientos a lo largo del curso de un estudio que tienen el potencial de afectar a la variable dependiente e introducir confusión en los resultados. Estos acontecimientos, que suelen deberse a factores ambientales, sociales o personales de los participantes, pueden controlarse mediante varios métodos, incluida la utilización de grupos de control y la implementación de técnicas de aleatorización, coherencia y eliminación para mantener la estabilidad en todos los grupos. Al garantizar que toda la muestra participe simultáneamente en el experimento en condiciones uniformes, la influencia de los acontecimientos históricos puede minimizarse o estandarizarse en todos los ámbitos. En algunos diseños de investigación, como los estudios cuasiexperimentales en los que los grupos están predeterminados antes de la selección, puede producirse un fenómeno conocido como historia local, que implica una

interacción entre la selección y los acontecimientos históricos. Esta amenaza surge de la posibilidad de que factores externos puedan afectar a diferentes grupos de manera desigual debido a las variaciones en los antecedentes de los sujetos (Shadish et al., 2002).

La maduración se refiere a los procesos internos que pueden desarrollarse dentro de los participantes con el tiempo, independientemente de cualquier tratamiento específico. Estos cambios, que pueden deberse a factores como la adaptación, la fatiga o el crecimiento biológico y psicológico, son particularmente frecuentes cuando transcurre un tiempo considerable entre la aplicación del tratamiento y la medición de los resultados, especialmente en el caso de los niños que experimentan transiciones rápidas en el desarrollo. A diferencia de las influencias históricas, los efectos de maduración son inherentes al individuo y no están provocados por circunstancias externas. Para mitigar esta amenaza, acortar la duración del estudio e incluir un grupo de control que no haya estado expuesto al tratamiento podría ayudar a diferenciar entre los efectos de la maduración y otras variables (Ato & Vallejo, 2015).

Durante la administración de la prueba, la familiaridad que se adquiere con la repetición de las pruebas puede provocar distorsiones en la respuesta de los sujetos, especialmente cuando se utilizan pruebas similares o se administra la misma prueba varias veces. Este fenómeno es particularmente relevante cuando se emplean versiones paralelas de una prueba. La amenaza suele surgir en los diseños interdisciplinarios cuando se emplean medidas previas y posteriores a la prueba, que implican la interacción entre la prueba previa y el tratamiento, y en los diseños intrasujetos, en relación con el efecto de orden, el error progresivo, los efectos residuales o de arrastre. Este problema puede solucionarse en los diseños interdisciplinarios omitiendo la medida previa al ensayo o utilizando grupos de control no tratados con mediciones previas y posteriores; y en los diseños intrasujetos, empleando metodologías de contrabalanceo. El alcance del impacto de esta amenaza podría haberse evaluado utilizando dos grupos con medidas previas y posteriores a la prueba, en lugar de utilizar uno solo, administrando el tratamiento a un solo grupo. Para investigar más a fondo la interacción entre la medida previa a la prueba y el tratamiento, se podría utilizar un diseño de Solomon con cuatro grupos de sujetos: dos grupos con la medida previa y posterior (un grupo recibe el tratamiento y el otro no) y dos grupos con la medida posterior al tratamiento únicamente (aplicar el tratamiento a uno de los dos grupos) (Delgado et al., 2011). A continuación se indica la representación simbólica del diseño de Solomon para cuatro grupos:

Grupo 1: R O₁ X O₂

Grupo 2: R O₁ - O₂

Grupo 3: R - X O₂

Grupo 4: R - - O₂

Donde R indica asignación aleatoria, O las observaciones (pretest y posttest) y X la aplicación del tratamiento. Esta representación permite visualizar de forma clara la estructura del diseño y el control de la posible interacción entre la medida previa y el tratamiento.

En el ámbito de la instrumentación, la amenaza se refiere al proceso de medición de la variable dependiente y es provocada por alteraciones que pueden producirse con el tiempo en los dispositivos o procedimientos de registro o medición. Es imperativo garantizar una medición precisa y coherente del comportamiento durante todo el estudio y en todos los participantes. Cuando se utilizan instrumentos de laboratorio, pueden producirse problemas de calibración o un ligero deterioro debido al paso del tiempo. Especialmente en el caso de la metodología observacional y cualitativa, en la que los observadores son humanos, esta amenaza puede manifestarse con mayor frecuencia, ya que las mediciones dependen de la percepción y el juicio de los observadores durante la grabación; además, también pueden producirse cambios en los observadores debido al aprendizaje, la fatiga, la monotonía, la falta de impulso, las distracciones, etc. (Anguera, 1990). Esta amenaza se puede mitigar mediante la capacitación de los observadores y la utilización de herramientas de medición estandarizadas, válidas y confiables que midan de manera consistente lo que pretenden medir cuando se usan varias veces en los mismos participantes y condiciones. En ocasiones, los instrumentos de medición pueden ser inexactos, lo que lleva a errores en la medición del comportamiento en el mismo grupo de sujetos. Estos errores, considerados aleatorios, podrían aumentar la varianza de los errores de la investigación, una situación que puede abordarse empleando grandes grupos de sujetos (Pascual et al., 1995).

En la selección diferencial, la amenaza se asocia con la formación de grupos. En cualquier estudio, los grupos deben ser equivalentes antes de aplicar el tratamiento en relación con la variable dependiente y/u otras variables temáticas vinculadas al objetivo de la investigación. Esta equivalencia garantiza que los resultados se deriven del efecto del tratamiento y no de las discrepancias iniciales entre los grupos en estas variables. Si no se forman grupos de forma aleatoria o si los grupos son demasiado pequeños para que el

azar actúe de manera adecuada, pueden producirse grupos no equivalentes. La amenaza del sesgo de selección es particularmente frecuente en los diseños cuasiexperimentales, ya que utilizan grupos preexistentes, a diferencia de los diseños experimentales que emplean la aleatorización, los bloques o el emparejamiento. En los diseños interdisciplinarios, garantizar la equivalencia de los grupos implica la utilización de técnicas de asignación aleatoria o de bloques aleatorios para la asignación de los participantes. Por el contrario, los diseños intrasujeto requieren un muestreo aleatorio de la población. Los diseños no experimentales, por otro lado, requieren muestras grandes y representativas seleccionadas mediante procedimientos aleatorios (Ato & Vallejo, 2015).

La amenaza de mortalidad experimental se refiere a la pérdida desigual de participantes durante un experimento y está estrechamente relacionada con el sesgo antes mencionado. La equivalencia inicial de los grupos en un estudio puede verse comprometida si algunos sujetos se retiran de un grupo, lo que justifica que el tamaño de los grupos sea lo suficientemente grande como para mantener la aleatorización (Martínez, 1995).

En la regresión estadística, la amenaza está relacionada con la selección de los sujetos. En ocasiones, los sujetos no se asignan al azar a los tratamientos, sino en función de los valores de una variable específica. Esta práctica suele provocar que los participantes con puntuaciones extremas converjan hacia la media en las mediciones posteriores debido a la falta de fiabilidad del instrumento de medición. Para mitigar esto, la selección de los sujetos debe ir precedida de varias mediciones. Además, este fenómeno puede confundir los diseños que incorporan medidas previas al tratamiento, ya que las puntuaciones extremas previas al tratamiento pueden normalizarse después del tratamiento, no debido al tratamiento en sí, sino a la variabilidad inherente de los datos (Martínez-Garza, 1988). Un ejemplo de este fenómeno puede observarse en diversos contextos. Por ejemplo, en un estudio educativo, si se selecciona a los estudiantes con las puntuaciones más bajas en matemáticas para aplicar un programa de refuerzo, es posible que en la evaluación posterior mejoren sus resultados. Sin embargo, esta mejora podría deberse parcialmente a la regresión a la media y no exclusivamente al efecto del programa. Del mismo modo, los estudiantes con puntuaciones inicialmente muy altas podrían mostrar un descenso en mediciones posteriores, sin que esto implique un empeoramiento real de su rendimiento. En el ámbito clínico, si se selecciona a pacientes con niveles extremadamente altos de ansiedad para aplicar un tratamiento, es probable que en evaluaciones posteriores sus puntuaciones disminuyan. No obstante, esta reducción puede estar influida por la

tendencia natural de las puntuaciones extremas a aproximarse a la media, independientemente de la eficacia del tratamiento. Asimismo, en el contexto deportivo, si se eligen atletas que han obtenido resultados excepcionalmente bajos en una competición para someterlos a un programa de entrenamiento específico, sus marcas posteriores pueden mejorar sin que necesariamente el programa sea la causa principal de dicho cambio.

6.3. Validez Externa

La validez externa se refiere a la capacidad de extrapolar los resultados de la investigación a diferentes poblaciones (validez poblacional), situaciones (validez ecológica) o períodos de tiempo distintos de los estudiados (validez histórica). Para que un experimento tenga validez externa, es fundamental que la muestra sea representativa de la población de interés y se haya seleccionado al azar. En esencia, «se refiere a la posibilidad de generalizar la conexión causal observada en un estudio determinado más allá de las condiciones en las que se estableció esa conexión» (Balluerka, 1999). Para garantizar la validez externa de un experimento, es crucial que la muestra sea representativa de la población objetivo y que se haya elegido al azar. Los principales desafíos para la validez externa surgen de cuestiones de interacción entre el tratamiento y varios aspectos del estudio (como la selección, la situación y el marco temporal), mientras que existen desafíos adicionales para la validez que son independientes de estas interacciones. Las amenazas relacionadas con la interacción entre el tratamiento pueden clasificarse en: interacción entre la selección y el tratamiento, interacción entre la situación y el tratamiento, e interacción entre el tiempo y el tratamiento. La primera desafía la validez poblacional, la segunda desafía la validez ecológica y la tercera desafía la validez histórica. Un estudio carece de validez poblacional cuando los resultados de una muestra no se pueden generalizar a la población objetivo u otras poblaciones porque la correlación establecida entre las variables es específica de los participantes del estudio. Dicho de otra manera, el tratamiento produce efectos específicos en un grupo de participantes debido a sus características únicas que no reflejan las de la población. Esto se debe a una interacción entre la selección y el tratamiento. Por ejemplo, un investigador quiere evaluar la eficacia de un programa de entrenamiento cognitivo diseñado para mejorar la memoria. Para ello, selecciona como muestra a un grupo de adultos mayores altamente activos intelectualmente (por ejemplo, personas que participan regularmente en actividades culturales, lectura y aprendizaje continuo). Tras la intervención, observa mejoras

significativas en el rendimiento de memoria. Sin embargo, estos resultados pueden no ser generalizables a la población general de adultos mayores, ya que el efecto del tratamiento podría estar influido por las características específicas de los participantes seleccionados (alto nivel de actividad cognitiva previa). En este caso, la mejora observada podría deberse a una interacción entre la selección (participantes con características particulares) y el tratamiento (programa cognitivo), lo que implica que el estudio carece de validez poblacional. Para mitigar esta amenaza, es esencial definir o restringir claramente las características de la población objetivo y seleccionar al azar la muestra de esa población. Si esto no es factible, es aconsejable intentar incluir una muestra de sujetos diversa en términos de variables relacionadas con el tratamiento y replicar el experimento con muestras variadas (Balluerka & Vergara, 2002). En el contexto de la interacción entre la situación y el tratamiento (situación $\times$ tratamiento), esta amenaza socava la validez ecológica. Se debe a la artificialidad, a veces excesiva, de los entornos experimentales, una situación que puede provocar que el contexto interactúe con el tratamiento y dificultar la generalización de los resultados a escenarios de la vida real. Esta amenaza prevalece en los métodos experimentales y es menos común en la investigación observacional y cualitativa. Para abordar este desafío, puede ser necesario replicar el estudio en diversos entornos y esforzarse por mantener la naturalidad en la situación experimental, dentro del rigor de investigación requerido. Si durante un estudio se producen acontecimientos que interactúan con el tratamiento, los resultados estarían directamente relacionados con ese momento específico y, por lo tanto, no podrían extrapolarse a diferentes momentos, lo que provocaría una deficiencia en la validez histórica del estudio. Esta amenaza se mitiga replicando el experimento en distintos momentos (Nosek, 2022).

Según Balluerka (1999), existen tres factores específicos que influyen en la validez: la interferencia de múltiples tratamientos, el impacto reactivo de las evaluaciones y el impacto reactivo de los instrumentos experimentales. La interferencia de varios tratamientos puede afectar a la validez de los resultados de la investigación. En los estudios en los que se administran varios tratamientos a los mismos sujetos, la respuesta a un tratamiento puede estar influenciada por los tratamientos anteriores, lo que lleva a una disminución de la validez interna y externa. En consecuencia, la capacidad de generalizar los resultados puede verse comprometida. Para abordar este problema, los investigadores pueden emplear la técnica de contrabalanceo, que consiste en variar sistemáticamente el orden de presentación de las condiciones experimentales entre los

participantes, de modo que todos los posibles órdenes se distribuyan equitativamente, controlando así los efectos de orden, la fatiga o los efectos de aprendizaje que podrían sesgar los resultados. Por otro lado, el efecto reactivo de las pruebas es similar a la interferencia de múltiples tratamientos antes mencionada y puede surgir en diseños en los que participan diferentes sujetos tras una evaluación previa a la prueba. Es importante señalar que la prueba previa puede sensibilizar a los sujetos, lo que no solo reduce la validez interna sino también la validez externa, lo que dificulta la generalización de los resultados. Por último, el efecto reactivo de los dispositivos experimentales es otro factor que puede afectar a los resultados de la investigación. El hecho de que los participantes sean conscientes de haber sido observados y evaluados durante un estudio puede llevarlos a formarse ideas o suposiciones sobre las expectativas de los experimentadores, lo que influiría en sus respuestas. Este fenómeno, conocido como características de la demanda, puede abordarse mediante la implementación de técnicas de simple ciego y doble ciego. Además, la novedad de la situación experimental también puede afectar a la validez interna y externa, particularmente en campos como la Psicología Clínica (Balluerka, 1999).

6.4. Validez de Conclusión Estadística

La validez de conclusión estadística se refiere esencialmente a dos inferencias estadísticas relacionadas que pueden afectar a la relación causa y efecto entre las variables (Lakens, 2022). En concreto si existe covarianza entre las variables en su causa y efecto, y cuál sería la magnitud empírica de su covariación. Para la primera de las inferencias se puede concluir de forma incorrecta que causa y efecto covarían cuando el hecho es que no covarían, lo cual produce lo que conocemos como Error Tipo I o también concluir incorrectamente que no covarían cuando el hecho es que sí lo hacen, en cuyo caso se cometería lo que conocemos como Error Tipo II (Ato & Vallejo, 2015). El método establecido para analizar la covariación de la causa y efecto es el contraste de hipótesis, en la cual se compara la hipótesis alternativa y la hipótesis nula mediante una prueba estadística para aceptar o rechazar la hipótesis nula. En el contraste de hipótesis cuando aceptamos la hipótesis nula siendo esta verdadera tomamos una decisión correcta, al igual que cuando rechazamos la hipótesis nula cuando la hipótesis nula es falsa. Sin embargo, cuando rechazamos la hipótesis nula siendo esta verdadera cometemos el Error Tipo I, mientras que cuando aceptamos la hipótesis nula siendo esta falsa cometemos el Error Tipo II (Delgado, 2011).

Existen varias amenazas contra la validez de conclusión estadística, situaciones en la que la inferencia estadística conlleva a Errores Tipo I y Tipo II o algún tipo de sesgo (Lakens, 2022). A continuación, exponemos algunas de las amenazas contra la validez de conclusión estadística. Una de las primeras amenazas a la validez de conclusión estadística es la baja potencia estadística, donde nos encontramos que una baja potencia estadística puede concluir de forma incorrecta la existencia de una relación no significativa (Cohen, 1988). Por otro lado, tenemos la violación de los supuestos de la prueba de estadísticas donde se puede infravalorar o sobrevalorar los tamaños y la significación de la efectividad del tratamiento. Por otro lado, tenemos lo que se conoce como violación de la tasa de Error Tipo I, pues se incrementa cuando aumenta el número de pruebas ejecutadas como, por ejemplo, al realizar varias comparaciones múltiples entre medias de tratamientos. Otra de las amenazas contra la validez de conclusión estadística sería la imprecisión de las medidas que ocurre cuando se utilizan medidas de baja fiabilidad lo que produce un aumento de la varianza error provocando conclusiones erróneas en la covariación entre variables. La baja fiabilidad en la aplicación de los tratamientos puede resultar amenazante para la validez de conclusiones estadísticas cuando los tratamientos no se aplican de la misma manera a todos los participantes lo cual aumenta la varianza error que conlleva a conclusiones erróneas sobre la convalidación. Por otro lado, tenemos la restricción del rango de la variable, que presentan restricciones en su variación tanto por el rango inferior como por el superior lo cual disminuye la potencia y debilita la inferencia sobre la convalidación. También puede resultar una amenaza contra la validez cuando hay presencia de varianza error en el contexto de la investigación, esto se debe a que algunas de las características del contexto conllevan a aumentar la varianza error que produce conclusiones erróneas en la covariación. Otra de las amenazas contra la validez de conclusiones estadísticas puede ser la existencia de una muestra muy heterogénea lo cual aumenta la desviación típica y disminuye la probabilidad de hallar la covariación entre variables. Por último, la estimación imprecisa de la magnitud del efecto puede conllevar a que se infravalore o se sobreestime lo cual provoque un error en la estimación de la covariación (Ato & Vallejo, 2015).

En resumen

- La validez es el grado en que una investigación mide lo que pretende medir y permite extraer las conclusiones que se pretenden extraer.

- Validez de constructo: ¿mide el instrumento el constructo teórico que pretende medir? Amenaza principal: instrumentos inadecuados o sesgos de respuesta.
- Validez interna: ¿son los cambios en la VD atribuibles a la VI y no a factores externos? Amenaza principal: variables extrañas no controladas (historia, maduración, selección).
- Validez de conclusión estadística: ¿son correctas las inferencias estadísticas? Amenaza principal: error Tipo I (falso positivo) y Tipo II (falso negativo), baja potencia estadística.
- Validez externa: ¿son generalizables los resultados a otras personas, contextos y momentos? Amenaza principal: muestras no representativas o condiciones artificiales de laboratorio.

7. Modelos de Análisis Estadístico

El análisis estadístico en Psicología experimental se organiza en este manual desde el enfoque del modelado estadístico unificado (Ato & Vallejo, 2015), que permite presentar los distintos procedimientos de análisis como variantes de un mismo marco conceptual. Para los objetivos de un manual de grado introductorio, el núcleo del análisis lo constituyen el Modelo Lineal (LM) y el Modelo Lineal Generalizado (GLM), que cubren la práctica totalidad de las situaciones de análisis que el estudiante encontrará en sus primeros años de investigación. Los modelos de mayor complejidad, como el Modelo Lineal Mixto (LMM), el Modelo Lineal Mixto Generalizado (GLMM), y los Modelos de Ecuaciones Estructurales (SEM), se presentan de forma resumida como extensiones del marco general, con el objetivo de que el estudiante conozca su existencia y su utilidad, pero sin pretender su dominio en este nivel formativo. Su estudio detallado corresponde a cursos de metodología avanzada.

La investigación que se lleva a cabo en el campo de las Ciencias del Comportamiento implica una amplia recopilación de datos, lo que lleva a intrincados procesos de tratamiento de datos que pueden provocar la omisión de información crucial relacionada con el tema investigado. Comprender la estructura o el marco de la información facilita la simplificación de la realidad empírica y, por lo tanto, la alinea más estrechamente con el modelo de estudio previsto. En el ámbito de la Psicología, se utilizan comúnmente modelos probabilísticos, que presentan variaciones aleatorias en las entradas y salidas del sistema. El manejo de esta aleatoriedad dentro de los modelos probabilísticos implica el empleo de una herramienta fundamental conocida como inferencia estadística (Ato & Vallejo, 2015; Lakens, 2022).

El enfoque predominante de la inferencia estadística en Psicología se basa en el contraste de hipótesis estadísticas mediante pruebas de significación acorde con la estadística clásica (Fisher, 1933). Las pruebas de significación de Fisher proponen formular la hipótesis nula (H_0) como verdadera y determinar la probabilidad condicional de los datos ante esta hipótesis [$p = (D/H_0)$] mediante un procedimiento estadístico. La probabilidad resultante (p) representa el nivel exacto de significación, denominado «valor p », lo que le valió al modelo de Fisher el sobrenombre de «enfoque del valor p » (Kline, 2004). Modificaciones posteriores del contraste de hipótesis de la estadística clásica introdujeron el concepto de la hipótesis alternativa (H_1) que representa las hipótesis que surgen cuando

se rechaza la H_0 (Neyman & Pearson, 1933). En este proceso de toma de decisiones, el investigador establece un nivel de significación (α), normalmente de 0,05, y lo compara con el valor p obtenido. El nivel de significación (α) indica que si H_0 es verdadero se rechazará erróneamente el 5% de las veces en experimentos repetidos. En consecuencia, surgen cuatro escenarios, dos que implican decisiones correctas y dos que implican decisiones erróneas. El Error de Tipo I se produce cuando se rechaza H_0 por error, con una probabilidad de α y un nivel de confianza complementario de $1-\alpha$. El Error de Tipo II se produce cuando se acepta incorrectamente H_0 , con una probabilidad beta y una potencia complementaria de la prueba $1-\alpha$. A consecuencia de numerosas críticas al enfoque de contraste de hipótesis relacionadas con discrepancias entre los objetivos de la investigación y la inferencia estadística, la improbable verificación de H_0 en caso de que no hubiera diferencias, la selección arbitraria de niveles de significancia de 0,05 y 0,01 y la incapacidad de los resultados estadísticamente significativos para evaluar la significación práctica de los hallazgos de la investigación (Kirk, 2013), se desarrolló un enfoque complementario de Comparación de Modelos. Centrada en el ajuste y la Comparación de los Modelos probabilísticos (Ato & Vallejo, 2015; Maxwell et al., 2008), el paso principal consiste en realizar un Modelado Estadístico, cuyo objetivo es proporcionar una descripción coherente del comportamiento de una variable dependiente (datos) a partir de una función ponderada de una o más variables independientes (modelo). Debido a la naturaleza inherente de los modelos probabilísticos, la explicación final no suele coincidir con precisión con los datos, por lo que es necesario evaluar la falta de ajuste entre los datos y el modelo establecido (residual o error). En notación matricial, el modelo puede expresarse como:

$$y_i = \mu_i + e_i$$

Donde y_i es denominado como el vector de valores observados que corresponde a una variable de respuesta, que depende del μ_i vector de valores estimados (componente sistemático) y un e_i vector residual (componente aleatorio). Esta ecuación representa las discrepancias existentes entre los valores observados y los valores estimados, lo cual constituye el componente aleatorio (Ato & Vallejo, 2015).

El objetivo principal del modelado estadístico es determinar cuál es el modelo más conciso capaz de explicar los datos con un mínimo de error (Ato & Vallejo, 2015). Se desarrolla en cuatro fases consecutivas para proporcionar la información adecuada y

determinar la idoneidad del modelo ajustado para la explicación o predicción de los datos. Sin embargo, antes de embarcarse en el proceso de modelización estadística, los investigadores deben considerar dos aspectos cruciales: el tipo de variable de respuesta y la terminología adecuada para distinguir los distintos modelos. Las variables de respuesta pueden ser numéricas (distribución normal), basadas en el recuento (distribución de Poisson), basadas en proporciones (distribución binomial), basadas en tasas (distribución de Poisson), ordinales (distribución multinomial) y binarias (distribución de Bernoulli). En cuanto a la terminología, se reconocen diferentes complejidades del modelo en función del número de parámetros: (a) modelo saturado, que abarca tantos parámetros como observaciones, que representa el modelo más complejo y menos conciso; (b) modelo nulo, que comprende un solo parámetro que denota el valor promedio de todas las observaciones, como el modelo más simple y conciso; (c) modelo óptimo o máximo, que incluye los parámetros de interés para el investigador, caracterizado por una discrepancia mínima en comparación con el modelo saturado o nulo. El proceso de modelización estadística implica la comparación de dos modelos: un modelo restringido con parámetros de interés específicos y un modelo extendido que contiene parámetros adicionales en relación con el modelo restringido. El objetivo es identificar el modelo que ofrece el mejor ajuste de datos, considerado como el modelo óptimo (Ato & Vallejo, 2015).

Como se ha comentado previamente, el proceso de modelado estadístico se desarrolla en cuatro fases con el propósito de identificar el modelo que mejor se ajusta con los datos. La etapa inicial, la especificación, implica proponer un modelo matemático basado en conocimientos sustantivos o evidencia empírica para predecir una variable de respuesta mediante predictores o variables explicativas. Los componentes clave incluyen el componente sistemático, cuyo objetivo es revelar el potencial de las variables explicativas para predecir la variable de respuesta, y el componente de error, que captura las variaciones aleatorias individuales mediante la selección de una distribución de probabilidad para la variable de respuesta. Entre las estructuras habituales de la modelización estadística figuran el Modelo Lineal (LM, Lineal Model), el Modelo Lineal Generalizado (GLM, Generalized Linear Model), el Modelo Lineal Mixto (LMM, Linear Mixed Model), el Modelo Lineal Mixto Generalizado (GLMM, Generalized Linear Mixed Model) y el Modelo de Ecuaciones Estructurales (SEM, Structural Equational Model). Para los objetivos de un manual de grado introductorio, el núcleo del análisis lo constituyen el Modelo Lineal (LM) y el Modelo Lineal Generalizado (GLM), que cubren

la práctica totalidad de las situaciones de análisis que el estudiante encontrará en sus primeros años de investigación. El Modelo Lineal (LM) permite analizar la relación entre una variable dependiente continua y una o varias variables independientes, asumiendo normalidad y relaciones lineales (incluye técnicas como la regresión lineal y el ANOVA). El Modelo Lineal Generalizado (GLM) amplía el modelo lineal al permitir variables dependientes que no siguen una distribución normal (por ejemplo, binarias o de conteo), mediante el uso de funciones de enlace y distribuciones como la binomial o la de Poisson. El Modelo Lineal Mixto (LMM) extiende el modelo lineal incorporando efectos aleatorios, lo que permite analizar datos con estructuras de dependencia, como medidas repetidas o datos agrupados (por ejemplo, estudiantes dentro de aulas). El Modelo Lineal Mixto Generalizado (GLMM) combina las características del GLM y del LMM, permitiendo trabajar con variables dependientes no normales y, al mismo tiempo, incluir efectos aleatorios. Por su parte, los Modelos de Ecuaciones Estructurales (SEM) permiten analizar relaciones complejas entre variables, incluyendo variables latentes no observadas directamente, así como relaciones simultáneas entre múltiples variables dependientes e independientes.

La etapa de ajuste, la segunda fase, implica estimar los parámetros del modelo mediante un procedimiento adecuado. Los mínimos cuadrados ordinarios (OLS, Ordinary Least Squares) se utilizan habitualmente en las estructuras LM para minimizar los cuadrados de error. Esta estructura asume una distribución de respuesta numérica normal y variable. Las estructuras GLM y GLMM suelen utilizar la máxima verosimilitud (ML, Maximun Likelihood) para la estimación de los parámetros, con el objetivo de maximizar la función de verosimilitud que representa las diferencias entre los valores observados y los estimados por el modelo. Independientemente de la estructura, la desviación (D) es una medida clave de la discrepancia entre los valores observados y estimados, definida como $D = -2LL$, donde LL (probabilidad logarítmica/Loglikelihood) es el logaritmo natural de la función de verosimilitud. Por lo general, los valores de desviación más altos indican un peor ajuste del modelo.

Hay dos procedimientos generales para el ajuste del modelo: el ajuste global y el ajuste condicional. La evaluación del ajuste global tiene como objetivo determinar si un modelo es adecuado para explicar los datos observados, aunque esta determinación no significa que sea el mejor modelo posible. En tales casos, es necesario realizar comparaciones entre modelos alternativos para determinar cuál es el modelo que mejor se ajusta a los datos.

Este proceso de comparación de modelos es aplicable únicamente a los modelos anidados (en los que un modelo es un subconjunto de otro y difiere en el número de parámetros). El ajuste condicional implica comparar modelos anidados, específicamente un modelo más simple (con menos parámetros) y un modelo más complejo (con más parámetros), mediante una prueba de razón de verosimilitud. Los estadísticos de la prueba siguen una distribución cuadrática χ^2 . Si las desviaciones entre los modelos muestran diferencias significativas, se prefiere el modelo más complejo; de lo contrario, se prefiere el modelo más simple. Los criterios de información como el AIC (Akaike Information Criterion) y el BIC (Bayesian Information Criterion) son beneficiosos en esta comparación, ya que los valores más bajos indican un mejor ajuste de los datos.

El tercer paso incluye la evaluación de las propiedades del modelo supuesto. Tanto los modelos LM como sus derivados requieren ciertas condiciones de estructura de datos, como la distribución normal de la variable de respuesta, la homogeneidad de las varianzas, la independencia de las observaciones y la linealidad entre las variables. Por último, la cuarta etapa consiste en interpretar el modelo en función de criterios lógicos, estadísticos y sustantivos.

7.1. Modelo Lineal Clásico

El Modelo Lineal Clásico (LM) se usa ampliamente en las Ciencias del Comportamiento e incluye procedimientos univariados como el análisis de regresión, el análisis de varianza (ANOVA), el análisis de covarianzas (ANCOVA), así como procedimientos multivariantes (MANOVA, MANCOVA y Análisis discriminantes). Se presenta la ecuación central en su forma univariada del LM,

$$Y_i = \beta_0 + \sum_{j=1}^p X_{ij}\beta_j + e_j$$

donde Y_i es la variable de respuesta, β_0 es el parámetro constante que representa el valor de la variable dependiente en el punto de referencia del modelo (cuando todas las variables explicativas son cero), el sumatorio de X_{ij} es el conjunto de valores observados de la p explicativas, β_j es el coeficiente de regresión asociado a cada variable explicativa, y e_i representa el error aleatorio. Esta formulación se puede representar en un sistema matricial como:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} X_{10} & X_{11} & \cdots & X_{1p} \\ X_{20} & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i0} & X_{i1} & \cdots & X_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N0} & X_{N1} & \cdots & X_{Np} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \\ \vdots \\ e_N \end{pmatrix}$$

Y cuya formula compacta de la ecuación con notación matricial es:

$$Y_{(Nx1)} = X_{(Nxq)}\beta_{(qx1)} + e_{(Nx1)}$$

Para la versión multivariante de m medidas de una variable de respuesta o m variables de respuestas diferentes, el sistema matricial multivariante de LM se representa como:

$$\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{i1} & Y_{i2} & \cdots & Y_{im} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{pmatrix} = \begin{pmatrix} X_{10} & X_{11} & \cdots & X_{1p} \\ X_{20} & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i0} & X_{i1} & \cdots & X_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n0} & X_{n1} & \cdots & X_{np} \end{pmatrix} \cdot \begin{pmatrix} \beta_{10} & \beta_{11} & \cdots & \beta_{1m} \\ \beta_{20} & \beta_{21} & \cdots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{i0} & \beta_{i1} & \cdots & \beta_{im} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p0} & \beta_{p1} & \cdots & \beta_{pm} \end{pmatrix} + \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ e_{i1} & e_{i2} & \cdots & e_{im} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix}$$

con notación matricial es:

$$Y_{(Nxm)} = X_{(Nxq)}B_{(qxm)} + E_{(Nxm)}$$

Donde Y representa la matriz de observaciones, X es la matriz del modelo, B es la matriz de parámetros, y E es la matriz de los errores. En la estructura univariante cuyos indicadores son vectores, en la estructura multivariante Y , B , y E son matrices.

La estructura univariada del LM tiene como características que la variable de respuesta Y es numérica, mientras que las variables explicativas X pueden ser categóricas y/o numéricas. La relación entre el valor esperado y el predictor lineal se mantiene en la misma escala, por lo que se utiliza una función de enlace de identidad. Se cumplen los

supuestos de normalidad de las respuestas de las variables, la homocedasticidad, la independencia y la linealidad.

7.2. Modelo Lineal Generalizado

El Modelo Lineal Generalizado (GLM) es una reformulación del modelo LM que combina los procedimientos LM con varios procedimientos específicos diseñados para variables de respuesta de diferente naturaleza, como las variables categóricas, de recuento y de frecuencia. Estos procedimientos incluyen el análisis logarítmico lineal, el análisis logit, la regresión logística y la regresión de Poisson, mientras que todos estos modelos comparten el cumplimiento de la linealidad y utilizan el procedimiento de ML para la estimación de los parámetros. La representación matemática del GLM utilizando la notación matricial es la siguiente:

$$Y_{(nx1)} = X_{(nxq)}\beta_{(qx1)} + e_{(nx1)}$$

donde Y representa los datos, $X\beta = \mu$ representa el modelo y e simboliza el error residual. Los atributos específicos del GLM comprenden una serie de características como es que la variable de respuesta Y puede adoptar diversas formas, como numérica, binaria, multinomial, ordinal, de recuento y de frecuencia. La matriz del modelo X puede abarcar una combinación de variables explicativas numéricas y categóricas. Es necesaria una función de enlace (g) para transformar el valor o modelo esperado (μ) y el predictor lineal (η) en la misma escala [$\eta = g(\mu) = X\beta$]. Los supuestos de los modelos de estructura GLM son menos estrictos que los de los modelos con estructura LM. Estos modelos pueden emplear cualquier distribución de la familia exponencial para la variable de respuesta, la varianza de la variable de respuesta puede no ser constante y la linealidad es esencial para la escala predictora, pero no para la escala de valores esperados. Sin embargo, la independencia es un supuesto crucial para todos los procedimientos de GLM (Zeger et al., 1988). La estimación de parámetros para los modelos con estructura GLM suele emplear procedimientos de máxima verosimilitud ML.

El GLM consta de tres componentes principales. El primero hace referencia al componente sistemático. Este componente se refiere al predictor lineal, que es una función de las variables explicativas y su estructura varía en función de la naturaleza de las variables explicativas, que pueden ser numéricas, categóricas o una combinación de ambas. Por otro lado, el componente aleatorio involucra el vector de la variable de respuesta aleatoria Y , que se caracteriza por una distribución de probabilidad conocida

con componentes independientes y un valor esperado μ . Las distribuciones de probabilidad pueden pertenecer a la familia exponencial, siendo las más comunes las distribuciones normal, binomial, multinomial y de Poisson. La elección de la distribución depende de la naturaleza de la variable dependiente: la distribución normal se utiliza cuando la variable es continua y aproximadamente simétrica; la distribución binomial cuando la variable es dicotómica (por ejemplo, éxito/fracaso); la multinomial cuando la variable presenta más de dos categorías nominales; y la de Poisson cuando la variable representa recuentos de ocurrencias en un intervalo determinado (por ejemplo, número de eventos). Por último, el componente de la función de enlace hace referencia a la función de transformación del predictor lineal en los valores esperados. En los GLM se utilizan diferentes funciones de enlace, los más comunes son las funciones de enlace con la identidad, en enlace logarítmico, y el enlace logit.

7.3. Modelo Lineal Mixto (Nivel avanzado)

Los modelos que se presentan a continuación (LMM, GLMM, SEM) superan el nivel formativo de un manual de grado introductorio. Se incluyen para que el estudiante conozca su existencia y su campo de aplicación. Para su estudio detallado véase Ato y Vallejo (2015) y Maxwell et al. (2018).

El Modelo Lineal Mixto (LMM) es un desarrollo de la estructura LM que postula la variable de respuesta como numérica y conforme a una distribución normal. A diferencia del marco GLM, la estructura LMM permite la inclusión de factores aleatorios, lo que resulta beneficioso cuando se infringen los supuestos de independencia y varianza constante. La estructura del LMM se representa mediante la siguiente formulación matemática (Laird & Ware, 1982):

$$Y_{(Nx1)} = X_{(Nxq)}\beta_{(qx1)} + Z_{(Nxt)}u_{(tx1)} + e_{(Nx1)}$$

donde Y simboliza el vector de datos, X indica la matriz de efectos fijos constantes del modelo, β representa el vector de efectos fijos, Z representa la matriz de efectos aleatorios del modelo, u significa el vector de parámetros de los efectos aleatorios, que se adhieren a una distribución normal, y e simboliza la matriz de errores residuales aleatorios. En particular, la única diferencia entre la estructura LM y la estructura LMM es la incorporación de efectos aleatorios adicionales en el modelo.

La estructura LMM es particularmente ventajosa para modelar escenarios en los que la variable de respuesta no presenta una varianza constante o carece de independencia, a pesar de asumir la normalidad con cierta flexibilidad. Estos casos suelen surgir en estudios en los que las unidades experimentales comparten rasgos parecidos, en los que las observaciones muestran estructuras jerárquicas que requieren la utilización de modelos de varios niveles, o en estudios de medición longitudinales o de medidas repetidas que implican múltiples evaluaciones de las mismas materias en distintos puntos temporales. En situaciones tan frecuentes en la investigación en Psicológica, la adopción de diseños con una estructura LMM resulta imprescindible para caracterizar eficazmente los datos (Ato & Vallejo, 2015).

La estimación de los parámetros de efectos fijos se logra mediante el procedimiento de mínimos cuadrados ordinarios generalizados. Los parámetros de efectos aleatorios y las matrices de covarianzas se determinan mediante métodos de estimación de ML, como la máxima verosimilitud restringida. Posteriormente, durante la fase de ajuste del modelo, se ejecuta un enfoque de comparación del modelo mediante pruebas de razón de verosimilitud de las varianzas (-2LL) y la consideración de los criterios de información (AIC y BIC). Por lo general, el modelo que muestra el ajuste óptimo se identifica con los valores más bajos de estos índices.

7.4. Modelo Lineal Mixto Generalizado (Nivel avanzado)

El Modelo Lineal Mixto Generalizado (GLMM) integra las estructuras del GLM y el LMM, lo que permite la utilización de variables de respuesta de diversa naturaleza, varias distribuciones de la familia exponencial y la incorporación de múltiples componentes aleatorios en el modelo. Este marco resulta particularmente ventajoso para el análisis de datos que muestran correlaciones con tipos de distribución diferentes, datos longitudinales afectados por el desgaste, datos incompletos, el desequilibrio o la dispersión de la muestra, además de ofrecer una solución para los requisitos de transformación de datos (Bandera & Pérez, 2018; Lo & Andrews, 2015). Dentro del marco GLMM, el predictor lineal se formula de la siguiente manera:

$$\eta = X\beta + Zu$$

donde X representa la matriz del modelo de efectos fijos, β significa un vector desconocido de efectos fijos, Z indica la matriz de efectos aleatorios del modelo y u

representa el vector de parámetros de los efectos aleatorios, que sigue una distribución normal. En el marco del GLMM, la estimación de los parámetros se lleva a cabo mediante la ML, lo que ayuda a mitigar el sesgo derivado de registros de datos incompletos, desequilibrados o espaciados temporalmente (Stroup, 2012). Este enfoque abarca variantes como la máxima verosimilitud restringida y la cuasiverosimilitud penalizada (Breslow & Clayton, 1993). Además, este marco emplea un algoritmo que facilita el cálculo del mejor estimador lineal imparcial de efectos fijos y del mejor predictor lineal imparcial de efectos aleatorios. Este último, al ser un estimador bayesiano, presenta el error cuadrático medio mínimo de la predicción en comparación con todos los predictores lineales imparciales. Además, estos modelos permiten examinar los datos correlacionados mediante efectos aleatorios y la estimación de los componentes de la varianza asociados tanto al modelo como al error residual. Las técnicas de estimación empleadas ayudan a minimizar los sesgos derivados de datos incompletos, desequilibrados o dispersos, así como a modelar las estructuras de error en los datos de medición longitudinales.

7.5. Modelo de Ecuaciones Estructurales (Nivel avanzado)

En la ecuación general de modelados dícese cuando los datos se obtienen mediante la estimación del modelo más el error, se requiere una ecuación única a través de la cual se define la variable de pendiente y se estructura el modelo. Dicho modelo se compone de varias variables independientes junto a sus distintos parámetros. Los Modelos de Ecuaciones Estructurales (SEM) representan una metodología versátil que permite modelar asociaciones intrincadas entre variables observables y variables latentes (Thurstone, 1947). Esta técnica combina elementos de los modelos de análisis factorial, conocidos como modelos de medición, y los modelos de regresión, denominados modelos estructurales. Los SEM se definen un método estadístico multivariante que se utiliza para estimar relaciones complejas entre variables observadas y latentes (Igolkina & Meshcheryakov, 2020). El principio subyacente de este enfoque implica la medición de las variables manifiestas mediante el análisis factorial, mientras que los modelos de error se emplean para abordar los errores de medición. En consecuencia, los modelos de regresión se utilizan para representar las verdaderas relaciones estructurales entre las variables sin errores de medición.

Dentro de los SEM y los modelos jerárquicos lineales, existe una lógica de investigación confirmatoria subyacente en la que los investigadores formulan un modelo, esencialmente

una hipótesis, para contrastarlo con los datos recopilados. Esto contrasta con los métodos de segmentación estadística que adoptan un enfoque exploratorio al identificar las variables predictoras que influyen en la variable de respuesta. Además, esta estrategia de modelización se distingue de los estudios experimentales tradicionales no solo por el diseño experimental sino también por la naturaleza de las variables. Los SEM tienen como objetivo descubrir las variables latentes como causas, derivadas de construcciones teóricas, mientras que los experimentos clásicos atribuyen los cambios a variables directamente observables. La metodología SEM sirve como un enfoque integral para cuantificar y evaluar las teorías sustantivas. Desde el punto de vista metodológico, los SEM abarcan una variedad de técnicas estadísticas que facilitan el establecimiento de relaciones entre las variables predictoras (continuas o discretas) y las variables de respuesta. Estas variables pueden medirse o estar latentes, y los modelos incorporan ambos tipos. Las variables latentes representan construcciones teóricas sin un procedimiento de observación directa, aunque sus manifestaciones pueden inferirse a través de medidas específicas de variables observables. Los SEM desempeñan un doble papel a la hora de evaluar la plausibilidad de hipótesis teóricas complejas al examinar las posibles relaciones entre los constructos y probar estas relaciones mediante medidas definitorias. Por lo tanto, los SEM consideran el error de medición presente en varias variables, una característica distintiva de las técnicas estadísticas multivariantes (Raykov & Marcoulides, 2006). Estos modelos suelen representarse mediante un diagrama que ilustra las interrelaciones entre las variables y un conjunto de ecuaciones que formalizan y articulan su función metodológica dentro del modelo.

Desde el punto de vista estadístico, las metodologías y los procesos analíticos de los SEM y los modelos multinivel se han documentado desde principios del siglo XX. El origen del SEM se remonta a 1934, cuando Wright introdujo modelos de análisis de rutas que se centraban en los tamaños relativos de las mediciones óseas. Los principios fundamentales de los modelos jerárquicos lineales han estado vigentes desde que Fisher (1925) introdujo el análisis de datos anidados en el contexto del análisis de la varianza.

En resumen

- El modelado estadístico busca el modelo más conciso que explique los datos con el mínimo error. El proceso tiene cuatro fases: especificación, ajuste, evaluación e interpretación.

- El Modelo Lineal General (LM) es el modelo básico: incluye la regresión, el ANOVA, el ANCOVA y el MANOVA. Asume normalidad, homocedasticidad, independencia y linealidad.
- El Modelo Lineal Generalizado (GLM) extiende el LM a variables de respuesta no normales (binaria, de recuento, ordinal) mediante una función de enlace. Cubre la regresión logística y de Poisson.
- Los modelos LMM, GLMM y SEM son extensiones avanzadas para datos con estructura jerárquica, medidas repetidas o variables latentes. Su uso requiere formación estadística especializada.

Conclusiones

En este capítulo se han presentado los fundamentos metodológicos y analíticos de la Psicología experimental desde una perspectiva integradora. A lo largo del texto, se han examinado los principales diseños de investigación, las amenazas a la validez, tanto interna como externa, y las estrategias para su control, destacando la importancia de una adecuada planificación del estudio para garantizar la calidad de las inferencias. Asimismo, se ha introducido el enfoque del modelado estadístico unificado como marco conceptual que permite comprender los distintos procedimientos de análisis como variantes de un mismo sistema, facilitando una visión más coherente y estructurada del análisis de datos. En este contexto, se han descrito los modelos fundamentales (LM y GLM) y sus extensiones más complejas (LMM, GLMM y SEM), atendiendo al nivel formativo del alumnado.

En conjunto, estos contenidos ponen de manifiesto que la investigación psicológica rigurosa requiere una adecuada integración entre diseño, medición y análisis, de modo que las conclusiones obtenidas reflejen de forma válida y generalizable los fenómenos estudiados. Este enfoque proporciona al estudiante una base sólida para avanzar hacia niveles más complejos de investigación y análisis en su formación metodológica.

Referencias

- Anguera, M. T. (2016). *Metodología de la investigación en Psicología*. Síntesis.
- Anguera, M. T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J., & Vallejo, G. (1995). *Métodos de investigación en Psicología*. Síntesis.
- Arnau, J. (1975). *Diseños experimentales en Psicología y Educación*. Trillas.
- Arnau, J. (1978). *Psicología experimental*. Ariel.
- Arnau, J. (1986). *Psicología experimental: Diseños y análisis de datos*. Trillas.
- Arnau, J. (1989). *Metodología de la investigación en Psicología*. Síntesis.
- Arnau, J., Anguera, M. T., & Gómez, J. (1990). *Metodología de la investigación en Ciencias del Comportamiento*. Murcia.
- Ato, M., & Vallejo, G. (2007). *Diseños de investigación en Psicología*. Pirámide.
- Ato, M., & Vallejo, G. (2015). *Diseños de investigación en Psicología*. Pirámide.
- Ato, M., López, J. J., & Benavente, A. (2013). A classification system for research designs in Psychology. *Anales de Psicología*, 29(3), 1038–1059.
- Bakeman, R., & Quera, V. (2023). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.
- Balluerka, N., & Vergara, A. I. (2002). *Diseños de investigación experimental en Psicología*. Prentice Hall.
- Bandalos, D. L., & Finney, S. J. (2018). *Factor analysis: Exploratory and confirmatory*. Guilford Press.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs*. Pergamon.
- Blanca, M. J. (2004). Diseños longitudinales en Psicología. *Psicothema*, 16(2), 234–241.
- Blanca, M. J., & Rando, B. (2005). Diseños factoriales incompletos. *Metodología de las Ciencias del Comportamiento*, 7(1), 25–41.
- Botella, J., & Sánchez-Meca, J. (2015). *Meta-análisis en Ciencias Sociales y de la Salud*. Síntesis.
- Broadus, R. N. (1987). Toward a definition of bibliometrics. *Scientometrics*, 12(5–6), 373–379.
- Campbell & Stanley (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.
- Campbell, D. T., & Stanley, J. C. (1979). *Diseños experimentales y cuasi-experimentales en la investigación social*. Amorrortu.
- Carey, S. S. (2011). *A beginner's guide to scientific method* (4th ed.). Wadsworth.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin.
- Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi-experimentation. En *Handbook of research methods*.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.
- Cross, R. M., et al. (1981). *Research in Psychology*. Harper & Row.
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis. *Journal of Business Research*, 133, 285–296.
- Fisher, R. A. (1935). *The design of experiments*. Oliver & Boyd.
- Gauch, H. G. (2012). *Scientific method in brief*. Cambridge University Press.
- Gómez-Luna, E., Fernando-Navas, D., Aponte-Mayor, G., & Betancourt-Buitrago, L. (2014). Metodología para la revisión bibliográfica. *Información Tecnológica*, 25(1), 25–34.
- González-Garay, A., Díaz-García, L., Chiharu, M., Anzo-Orsorio, A., & García de la Puente, S. (2018). Generalidades de los estudios de casos y controles. *Acta pediátrica de México*, 39(1), 72-80.
- Haig, B. D. (2014). *Investigating the psychological world*. MIT Press.
- Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación*. McGraw-Hill.
- Kazdin, A. E. (1992). *Single-case research designs*. Oxford University Press.
- Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137–152.
- Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). SAGE.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.
- Lakens, D. (2022). *Improving your statistical inferences*. PsyArXiv.
<https://doi.org/10.31234/osf.io/yxfvb>
- León, O. G., & Montero, I. (2004). *Métodos de investigación en Psicología y Educación*. McGraw-Hill.

- Levitt, H. M., et al. (2018). Journal article reporting standards for qualitative and mixed methods research. *American Psychologist*, 73(1), 26–46.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing experiments and analyzing data*. Routledge.
- Nosek, B. A. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719–748.
- Privitera, G. J. (2022). *Research methods for the behavioral sciences* (3rd ed.). SAGE.
- Rother, E. T. (2007). Systematic literature review. *Acta Paulista de Enfermagem*, 20(2), v–vi.
- Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.
- Sánchez-Meca, J. (2010). Meta-analysis in Psychological research. *Psychothema*, 22(3), 451–457.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, 22(11), 1359–1366.
- Skinner, B. F. (1956). A case history in scientific method. *American Psychologist*, 11(5), 221–233.
- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, 46(2), 137–150.
- Taylor, S. J., & Bogdan, R. (1987). *Introducción a los métodos cualitativos de investigación*. Paidós.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Univ. of Chicago Press.
- Tversky, A., & Kahneman, D. (1973). Availability heuristic. *Cognitive Psychology*, 5(2), 207–232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty. *Science*, 185(4157), 1124–1131.
- Valsiner, J., et al. (2017). *Methodology in psychology*. Springer.
- Vautier, S. (2011). The operationalization of general hypotheses versus the discovery of empirical laws. *Philosophia Scientiae*, 15(1), 105–130.
- Vázquez, F. (2021). *Historia del pensamiento científico*. Síntesis.