

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/344237456>

Diseño de la investigación, análisis y redacción de los resultados (Edición 1ª, septiembre de 2020, Valencia)

Book · September 2020

DOI: 10.1.17605/OSF.IO/KNGTP

CITATIONS

0

READS

221

2 authors:



Dolores Frías-navarro
University of Valencia

132 PUBLICATIONS 2,016 CITATIONS

[SEE PROFILE](#)



Marcos Pascual-Soler
ESIC Business & Marketing School

43 PUBLICATIONS 254 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Impacto económico y social del emprendimiento y los nuevos modelos de negocio [View project](#)



Children 's education on the Internet: reducing risks and increasing benefits [View project](#)

Diseño de la investigación, análisis y redacción de los resultados

Editores:

Dolores Frías-Navarro

Marcos Pascual-Soler



**Edición 1ª, septiembre de 2020
Valencia**

Diseño de la investigación, análisis y redacción de los resultados

Disseny de la investigació, anàlisi i redacció dels resultats
Research design, analysis and writing of results

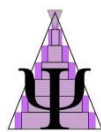
Editores:

Dolores Frías-Navarro

Marcos Pascual-Soler

*Universidad de Valencia, España

** ESIC Business & Marketing School, España



Edición 1ª, septiembre de 2020

DOI: 10.17605/OSF.IO/KNGTP

Edición 1ª, septiembre de 2020

España: Valencia

Portada y contraportada diseñada por Dolores Frías Navarro

Las fotografías pertenecen al álbum particular de Dolores Frías Navarro

Este material es un libro gratuito, depositado en abierto y en constante actualización. Las actualizaciones de su contenido que se produzcan durante el curso académico vigente se incorporarán en una nueva edición con los cambios efectuados y se anotará su versión y año.

La solución a los ejercicios de auto-evaluación (Anexo) se difundirá a través de la página Web:
<https://www.uv.es/friasnav/libroDiseño>

Citar el libro como:

Frías-Navarro, D. y Pascual-Soler, M. (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP>
Disponible en: osf.io/kngtp

@ Grupo de Investigación UV: GIUV2018-427. Universidad de Valencia

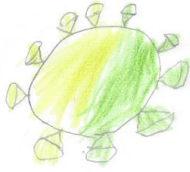
Laboratory: REsearch MEthods and design in applied psychology (**REME**)

<https://www.uv.es/friasnav/>

Comentarios y anotaciones en: M.Dolores.Frias@uv.es

*A las alumnas y alumnos
que inician su curso académico 2020/2021,
marcado para ser histórico
por la presencia del coronavirus y la COVID-19*

PRESENTACIÓN



14 de marzo de 2020, el Gobierno español decreta el estado de alarma en todo el territorio nacional durante 15 días. El Congreso de los Diputados autorizó al Gobierno a prorrogar el estado de alarma en seis ocasiones. La medida del estado de alarma se extendió hasta el 21 de junio. Hubo ciertas concesiones a partir de finales de abril. Así, desde el 26 de abril se permitió la salida de menores de catorce años acompañados por un adulto durante una hora, pero solamente en las inmediaciones de sus domicilios. La medida se amplió a los mayores de dicha edad desde el 2 de mayo, solo para pasear o hacer deporte de manera individual. El 21 de mayo se hizo obligatorio el uso de mascarillas en espacios públicos a partir de los 6 años. Se trataba de una pandemia mundial provocada por un nuevo coronavirus (CoV) conocido como SARS-CoV-2 y la enfermedad que causa se conoce por la COVID-19 (acrónimo en inglés de "COronaVirus Disease").

El resto de la historia ya la hemos vivido todos los españoles y el mundo en general. La aparición de la COVID-19 y la declaración de pandemia por la Organización Mundial de la Salud (OMS, 11 de marzo de 2020) fue muy rápido. Fue tan rápido que se tomaron medidas drásticas en muy pocos días. Tan drásticas como el estado de alarma nacional que supuso el confinamiento de la población en sus hogares. La libre circulación de los ciudadanos y ciudadanas se limitó a unas circunstancias muy concretas como adquirir alimentos y medicamentos, acudir a un centro médico o acudir al lugar de trabajo en unos puestos muy específicos y siguiendo unas rigurosas normas de seguridad. Y, por supuesto, los centros escolares y las universidades se cerraron. Después de unas semanas para afrontar el grave problema que supuso el coronavirus, comenzaron a tomarse medidas como pasar a una enseñanza virtual (*on-line*). Los profesores y las profesoras comenzaron a impartir docencia utilizando Internet, desde sus casas y con los medios personales que disponían. Se hizo todo lo posible por acabar el curso académico 2019/2020 en las mejores condiciones. Nadie podía prever al comienzo del curso, en septiembre de 2019, esta situación. Situación que ya es histórica.

Han pasado seis meses desde el estado de alarma, y la situación sanitaria sigue siendo muy crítica. Se ha pasado a lo que se denomina "la nueva normalidad". La distancia social, el uso de mascarillas y de gel hidroalcohólico se han convertido en conductas

obligatorias. Las aperturas del curso académico 2020/2021 en todas las universidades españolas se ha caracterizado por planificar dichas medidas durante el acto institucional. Una imagen que nunca pudimos imaginar. Las directrices sobre las medidas de actuación y prevención para el profesorado y alumnado durante el primer cuatrimestre del curso siguen



describiendo una situación de alarma sanitaria. Por supuesto, si se siente algún tipo de sintomatología relacionado con la COVID-19 no acudir al centro e iniciar el aislamiento social y hacerse la prueba PCR. Y, la mascarilla siempre puesta. Distancia social de 2 metros y permanecer en la facultad el menor tiempo posible. La docencia, en gran parte de las

facultades, será semi-presencial con clases presenciales y clases *on-line* sincrónicas. Nada que ver con el inicio tradicional del curso académico, con reencuentros, saludos y abrazos y muchas cosas que contar en los bancos de los pasillos y en las mesas de la cafetería. El curso académico 2020/2021 que ahora comienza no sabemos en qué condiciones acabará, ni siquiera sabemos cómo acabará el trimestre. Quizás la Universidad cierre sus puertas, quizás sea necesario pasar a la docencia totalmente *on-line* si las circunstancias sanitarias lo exigen. No lo sabemos. Ahora tenemos la experiencia anterior, pero la incertidumbre permanece debido a los datos sanitarios.



Ante esta situación decidimos ponernos a trabajar en la elaboración del material de este libro, "*Diseño de la investigación, análisis y redacción de los resultados*". Su objetivo es recoger en un formato libro los principales apuntes de la materia de diseños de investigación y permitir que el alumnado pueda consultar y estudiar los aspectos más relevantes para llevar a cabo todo el proceso del diseño de una investigación. Se trata de

un material de docencia dirigido a la formación de alumnos y alumnas principalmente de Psicología. El libro está depositado en un repositorio en abierto y, por lo tanto, el acceso y su descarga es gratuito. Además, decidimos que debía ser un material en constante actualización, añadiendo temas y reflexiones necesarias para la formación en los contenidos del área de la metodología de investigación. En esta primera edición se han elaborado los contenidos básicos del diseño de investigación.

La elaboración de esta primera edición del libro "*Diseño de la investigación, análisis y redacción de los resultados*" permite que el alumnado acceda a una explicación detallada de los conceptos y técnicas que se van a desarrollar en la materia de diseños de investigación. Este material, junto con las diapositivas de las clases y las lecturas de otros materiales recomendados por el profesorado, permitirá que puedan llevar a cabo un aprendizaje activo. Este tipo de aprendizaje implica estudiar de forma autónoma y acceder a las clases con cierto nivel de conocimiento, favoreciendo dinámicas de trabajo en clase dirigidas a potenciar la solución de problemas y avanzar en la formación. Con ello, se evita la dinámica de clase donde el alumnado es un receptor de información sincrónica, pasivo y solo puede comenzar su formación una vez ha recibido la clase.

El contenido del libro "*Diseño de la investigación, análisis y redacción de los resultados*" se ha planificado en cinco bloques o partes: 1) Elementos del proceso de diseño de la investigación, 2) Diseños de investigación, 3) Herramientas de trabajo de consulta del conocimiento previo (bases de datos) y manejo de programas estadísticos, 4) Lectura crítica de investigaciones y análisis de la presentación y redacción de resultados, 5) Ejemplos de investigaciones. Análisis de su estructura y lectura crítica.

La primera edición del material (septiembre de 2020) recoge el primer bloque (Elementos del proceso de diseño de la investigación) y del bloque dos (Diseños de investigación) se ha redactado el tema del diseño entre-grupos unifactorial univariado junto con las pruebas de contraste de hipótesis específicas. El resto de contenidos se irán preparando y formarán parte del libro en la siguiente actualización del material. Durante este curso académico, el alumnado recibirá la información necesaria para completar el temario a través de apuntes y materiales recomendados por el profesorado.

Desde el área de la metodología de investigación y la calidad de los resultados científicos, el comienzo del siglo XXI supuso reabrir de nuevo el debate sobre la replicación de los resultados, su reproducibilidad, el sesgo de publicación, los falsos positivos y los usos y abusos de las pruebas tradicionales de significación estadística junto con la necesidad de la educación y la re-educación estadística. Los elementos que rodean a la actual cultura de la investigación, especialmente la necesidad de publicar y publicar en

revistas de impacto, ha conducido a un escenario que no favorece la búsqueda de la mejor evidencia (prueba) y, en algunos casos, ha sesgado las actividades del científico o científica. Por ello, la competencia de la lectura crítica o activa se ha convertido en esencial para el investigador o investigadora y para el lector o lectora de literatura científica. Dicha lectura requiere aproximarse a los resultados científicos desde una perspectiva de análisis metodológico y con cierto grado de escepticismo que mantenga activa la mirada crítica. La formación en los contenidos del área de la metodología de investigación ha sido siempre necesaria para el científico o científica que lleva a cabo estudios empíricos y para los lectores y lectoras de ese tipo de información. Y, actualmente, sigue siendo esencial, primero para poder seleccionar, de la gran cantidad de información que se publica, los artículos o informes que aportan resultados con las mejores pruebas o evidencias y, sobre todo, porque se ha demostrado que los investigadores e investigadoras hacen, en ocasiones, un uso abusivo y sesgado de sus grados de libertad fomentando el sesgo de publicación que tantos problemas ha creado a la Ciencia y a la interpretación de la realidad de los fenómenos. La revolución de la denominada Ciencia en Abierto (“Open Science”), junto con otras conductas relacionadas como el pre-registro de los protocolos de investigación, el depósito de los datos obtenidos en el estudio y sus códigos y el fomento y consideración de los trabajos de meta-investigación, podría ayudar al cambio en la cultura de la investigación, a la forma de hacer Ciencia y a la forma de comunicar los resultados. Sin embargo, para cambiar la conducta de los investigadores y las investigadoras es necesario, en primer lugar, cambiar esa cultura y el sistema de incentivos académicos dirigidos a la promoción académica y profesional. Y, en segundo lugar, fomentar desde las aulas (ya desde las primeras etapas de la formación escolar) la educación en el razonamiento científico, en el método científico y en la elaboración y uso de hipótesis científicas. Por supuesto, la comunidad científica debe apostar fuerte por la ciencia básica y la elaboración de teorías que expliquen y den sentido a los hallazgos.

En definitiva, en el contexto actual de la Ciencia, la formación en los contenidos del área de la metodología de investigación (análisis de datos, psicometría y diseños de investigación) aporta unas herramientas fundamentales para acercarse a la elaboración de estudios y a la lectura de sus resultados con una actitud activa de rastreo y análisis crítico. Y, buscando el desarrollo de esas dos competencias se ha elaborado el material del libro “*Diseño de la investigación, análisis y redacción de los resultados*”.

Dolores Frías Navarro

Marcos Pascual Soler

11 de septiembre de 2020

ÍNDICE

PARTE I. Elementos del proceso de diseño de la investigación

CAPÍTULO 1. Investigación científica en Psicología y diseño del

estudio (Dolores Frías-Navarro) 15

| | |
|--|----|
| Aspectos éticos en el manejo de la investigación psicológica | 18 |
| Grados de libertad del investigador o investigadora | 20 |
| Método científico | 24 |
| Reforma estadística y Práctica Basada en la Evidencia | 29 |
| Diseño de investigación..... | 32 |

CAPÍTULO 2. Variables del estudio (Dolores Frías-Navarro) 37

| | |
|---|----|
| Clasificación de las variables del estudio | 37 |
| Criterio metodológico | 38 |
| Variable independiente..... | 39 |
| Variable independiente: manipulada / asignada | 40 |
| Variable dependiente | 42 |
| Variable extraña | 42 |
| Criterio estadístico..... | 46 |

CAPÍTULO 3. Proceso del diseño de investigación

(Dolores Frías-Navarro) 49

| | |
|---|----|
| Necesidad de conocimiento | 49 |
| Conocimiento previo..... | 52 |
| Hipótesis de investigación | 61 |
| Planificación de la investigación | 61 |
| Método | 64 |
| Análisis de los datos..... | 64 |

| | |
|-----------------------------|----|
| Conocimiento adquirido..... | 65 |
|-----------------------------|----|

CAPÍTULO 4. Metodologías de investigación (Dolores Frías-Navarro) 69

| | |
|--|----|
| Diferencias entre las metodologías de investigación | 71 |
| Metodologías: experimental, cuasi-experimental y no experimental | 73 |
| Otras clasificaciones de las metodologías de investigación..... | 79 |
| Proceso de diseño de un estudio con metodología experimental | 81 |
| Metodología cuasi-experimental..... | 89 |
| Metodología no experimental | 90 |
| Asignación aleatoria del tratamiento..... | 92 |
| Diseño de $N = 1$ | 97 |

CAPÍTULO 5. Validez de los resultados de la investigación (Marcos Pascual-Soler y Dolores Frías-Navarro) 103

| | |
|--|-----|
| Validez interna | 107 |
| Validez de conclusión estadística..... | 111 |
| Validez de constructo | 114 |
| Validez externa | 116 |

PARTE II. Diseños de investigación

CAPÍTULO 6. Resultados de la investigación (1). Diseño entre-grupos, unifactorial, univariado (Dolores Frías-Navarro y Marcos Pascual-Soler) 119

| | |
|---|-----|
| Hipótesis científica, hipótesis estadística..... | 120 |
| El Factor Bayes (BF) | 127 |
| Resultados nulos..... | 129 |
| Ecuación estructural del modelo..... | 131 |
| Diseño de grupos independientes unifactorial univariado | 132 |
| Análisis de la varianza (ANOVA) | 135 |
| Modelos de la hipótesis nula y la hipótesis alternativa..... | 136 |

| | |
|---|-----|
| Estimación de los parámetros: término del efecto de A | 140 |
| Estimación de los parámetros: término de error | 140 |
| Razón F | 142 |
| Suma de Cuadrados | 142 |
| Grados de Libertad y Medias Cuadráticas..... | 143 |
| Decisión estadística | 144 |
| Errores estadísticos..... | 146 |
| Supuestos estadísticos del modelo | 149 |
| Desarrollo de la ecuación estructural. Ejemplo 1 | 150 |
| Grados de libertad totales..... | 158 |
| Grados de libertad del efecto..... | 158 |
| Grados de libertad del error..... | 159 |
| Redacción de resultados | 159 |
| La proporción de varianza explicada: R^2 , η^2 y η^2 parcial..... | 161 |
| R^2 y η^2 | 161 |
| Diseños factoriales: η^2 y η^2 parcial | 162 |
| Eta Cuadrado: η^2 | 162 |
| Eta Cuadrado parcial: η_p^2 | 163 |
| Comprobación de hipótesis específicas (diseño entre grupos $A < 2$)..... | 165 |
| Tasa de error de tipo I | 173 |
| Pruebas de contraste de hipótesis específicas | 175 |
| Procedimiento DHS (Honestly Significant Difference) de Tukey | 176 |
| Procedimiento de Dunnett | 177 |
| Corrección de Bonferroni..... | 177 |
| Procedimiento de Scheffé..... | 178 |
| Análisis con el programa SPSS..... | 185 |
| SPSS: ANOVA de un factor para muestras o grupos independientes..... | 187 |
| Análisis con el programa JASP | 189 |
| Referencias | 193 |

Anexos

Anexo 1. Breve explicación de conceptos fundamentales de diseño de investigación 199

| | |
|---|-----|
| Diseño de entre-grupos o entre-sujetos..... | 199 |
| Diseño de medidas repetidas o diseño intra-sujetos..... | 199 |
| Diseño mixto o de medidas parcialmente repetidas..... | 200 |
| Diseño de bloques..... | 200 |
| Diseño con variables covariadas, ANCOVA | 201 |
| Diseño de grupo equivalente o diseño con grupo de control equivalente | 202 |
| Diseño de grupo no equivalente o diseño con grupo de control no equivalent..... | 202 |
| Análisis de la Varianza | 202 |
| Valor p | 203 |
| Tamaño del efecto..... | 204 |
| Error de Tipo I | 205 |
| Nivel de confianza | 205 |
| Error de Tipo II | 205 |
| Potencia estadística | 205 |
| Revisión sistemática y meta-análisis | 206 |
| Referencias..... | 207 |

Anexo 2. Plantilla de aprendizaje: descomposición de la ecuación estructural. Diseño entre-grupos 209

Anexo 3. Tablas estadísticas 209

| | |
|--|-----|
| On-line: Calcular el valor teórico de la Razón F: http://davidmlane.com/hyperstat/F_table.html | 211 |
| Distribución F , alfa = .05..... | 211 |
| Distribución F , alfa = .01..... | 212 |

| | |
|--|-----|
| Distribución F , $\alpha = .001$ | 213 |
| Distribución t de Student | 214 |

Anexo 4. . Comparación entre diferentes pruebas: tamaño del efecto pequeño, mediano y grande 215

| | |
|--|-----|
| On-line: Calcular el tamaño del efecto y su intervalo de confianza: https://campbellcollaboration.org/research-resources/effect-size-calculator.html | 215 |
| Pruebas estadísticas y valores de los tamaños del efecto | 215 |

Anexo 5. Autoevaluaciones 217

| | |
|---------------------------|-----|
| 1) Autoevaluación 1 | 217 |
| 2) Autoevaluación 2 | 220 |

Otros capítulos (*en preparación*):

PARTE II. Diseños de investigación

- Procedimiento clásico de inferencia estadística. Perspectiva frecuentista
- Resultados de la investigación (2). Diseño entre-grupos, factorial, univariado Diseño factorial
- Resultados de la investigación (3). Diseño de bloques
- Resultados de la investigación (4). Diseño de medidas repetidas
- Resultados de la investigación (5). Diseño mixto o de medidas parcialmente repetidas
- Resultados de la investigación (6). Diseño con variables covariadas (ANCOVA)
- Resultados de la investigación (7). Diseños multivariados

PARTE III. Herramientas de trabajo de consulta del conocimiento previo (bases de datos) y manejo de programas estadísticos

- Consulta de base de datos. Búsqueda del conocimiento previo
- Uso básico del SPSS

-Análisis paralelos frecuentista-bayesianos con JASP. José Perezgonzalez

PARTE IV. Lectura crítica de investigaciones y análisis de la presentación y redacción de los resultados

-La Ciencia que se hace en el siglo XXI. Cultura de la investigación. Prácticas de investigación cuestionables. El fraude científico. Los debates sobre el método científico

-Recomendaciones para elaborar el informe de investigación. Formato del Manual APA (7ª edición)

-Lectura crítica. Análisis de artículos publicados

PARTE V. Ejemplos de investigaciones. Análisis de su estructura y lectura crítica

-Análisis de la estructura factorial de la Escala de Arraigo de inmigrantes latinoamericanos en España. José Berríos-Riquelme, Viviana Vargas-Salinas y Manuel Martín-Fernández

Capítulo I. Investigación científica en Psicología y diseño del estudio

Dolores Frías-Navarro
Universidad de Valencia

Citar el capítulo como:

Frías-Navarro, D. (2020). Investigación científica en Psicología y diseño del estudio. En D. Frías-Navarro y M. Pascual-Soler (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España.
DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> Disponible en: osf.io/kngtp

Mejorar la práctica estadística y la educación estadística es algo más que una necesidad urgente para todo el estudiantado universitario de cualquier disciplina científica y especialmente para el alumnado de Ciencias Sociales y de la Salud. Gran parte del cuerpo de conocimiento de estas disciplinas, por ejemplo la Psicología o la Medicina, se alcanza con la aplicación de lo que genéricamente se llama el ‘método científico’ donde la Estadística, el Diseño de la Investigación, la Psicometría y la Matemática forman parte inexcusable.

La necesidad de información y de formación constituyen dos condiciones básicas de la práctica del profesional de la Psicología, la Medicina, la Psiquiatría, la Sociología, el Trabajo Social, la Economía, la Biología, el Derecho, la Veterinaria, la Farmacia, la Enfermería... La actualización constante de los conocimientos adquiridos tras la formación académica es un requisito imprescindible para realizar la labor profesional con el mayor grado de éxito y calidad, asegurando de este modo que no haya deterioro del saber o conocimiento y, como consecuencia, también del ejercicio profesional.

Si se reflexiona sobre las características de un buen profesional, se podría llegar a la conclusión de que ser un buen profesional es cada vez más difícil dado el gran número de tareas que implica y la constante actualización que requiere. Estamos en un momento de gran evolución del conocimiento científico que además, avanza muy rápidamente. Todo ello exige estar actualizados en los avances que se producen en el campo de conocimiento del profesional donde la elaboración y publicación de investigaciones aplicadas y/o empíricas supone de forma inexcusable un referente básico que hay que consultar y valorar desde una perspectiva crítica o activa. Actualmente para ser un buen profesional ya no basta con “tener una gran experiencia” o una gran antigüedad en su campo de trabajo. Ahora es necesario formarse de forma continua (‘aprendizaje a lo largo de la vida’) y conocer los avances que se producen (diariamente), aplicando criterios de valoración crítica o lectura activa de la calidad de las pruebas empíricas o hallazgos (‘evidencia’) que se consultan.

Desde un modelo de Práctica Basada en la Evidencia que impera en la Ciencia del siglo XXI (la denominada Ciencia Basada en la Evidencia), la autoridad de los años y de la supuesta experiencia del profesional es sustituida por la autoridad de la investigación basada en la mejor evidencia o en las mejores pruebas (investigación realizada con el diseño de investigación que aporte los resultados más válidos y, por lo tanto, más próximos a la realidad del fenómeno estudiado). Y para obtener pruebas empíricas de calidad (resultados o evidencia) es necesario dominar de forma adecuada todos los elementos implicados en el proceso de diseño de investigación.

Los elementos o herramientas que forman parte del diseño de una investigación incluyen cuestiones como necesidad de conocimiento (problema de investigación), constructos, variables, tipos de hipótesis, sesgo, control, estadística, inferencia, estadísticos, contraste de hipótesis, resultados, redacción, conclusiones..., cuya presencia en el diseño está dirigida a dar solución al problema de investigación planteado. Se podrían comparar dichos elementos, por ejemplo, con las herramientas que los cirujanos utilizan para llevar cabo con éxito una operación (Imagen 1).

La lección de anatomía del Dr. Nicolaes Tulp

Rembrandt, 1632



COVID-19

Todos los elementos (herramientas) del diseño de la investigación:

Necesidad de conocimiento (problema de investigación), constructos, variables, tipos de hipótesis, sesgo, control, estadística, inferencia, estadísticos, contraste de hipótesis, resultados, redacción, conclusiones



Imagen 1. Elementos necesarios para ejecutar el diseño del estudio.

Imagen disponible en:

<https://ih1.redbubble.net/image.1179174666.1308/mp,504x498,matte,f8f8f8,t-pad,600x600,f8f8f8.jpg>

Y el modelo de la Práctica Basada en la Evidencia no sólo se aplica para obtener pruebas válidas con la elaboración y ejecución de una investigación propia, pues dicho modelo de actuación también se aplica cuando el lector o lectora hace una lectura 'crítica' o activa de los hallazgos que otros investigadores o investigadoras han publicado. En ambas situaciones (como investigador o investigadora y como lector o lectora) es necesario dominar los contenidos del ámbito de la metodología de investigación. En otras palabras, valorar un resultado de un estudio sólo puede ser realizado si el lector o lectora chequea un conjunto de elementos relacionados con el proceso del diseño de investigación que podrían invalidar los resultados del estudio (amenazas a la validez de los resultados) o, por el contrario, podrían otorgar calidad a los hallazgos o pruebas (validez) y de este modo sí aportarían evidencia válida al campo de trabajo del que forma parte el estudio.

El psicólogo o la psicóloga (el especialista o el profesional en general) es el responsable de su formación y debe garantizar que está preparado profesionalmente para abordar la intervención o el estudio de las variables con la mayor calidad disponible y, así, se evitan errores en las decisiones e intervenciones clínicas. La valoración de las propiedades o calidad de los hallazgos (lectura crítica o activa de los resultados de la investigación) debe realizarla desde una perspectiva científica apoyada en la calidad de la evidencia o las pruebas encontradas.

Aspectos éticos en el manejo de la investigación psicológica

En primer lugar, los profesionales de la Salud y las Ciencias Sociales se encuentran en un continuo desarrollo profesional y la responsabilidad ética exige la actualización constante de los conocimientos que garantice la justificación de las decisiones adoptadas dentro de un modelo basado en las mejores pruebas, resultados o evidencia disponible (modelo de Práctica Basada en la Evidencia). La relación entre los años que han transcurrido desde la graduación y los conocimientos actualizados que tiene el profesional sobre el mejor tratamiento es estadísticamente significativa y negativa (“curva resbaladiza” de la evolución del conocimiento tras la graduación). De ahí la necesidad de una formación continua y actualizada a lo largo de la trayectoria profesional.

En España el Código Deontológico del Psicólogo (2010) señala que “la autoridad profesional del Psicólogo/a se fundamenta en su capacitación y cualificación para las tareas que desempeña. El/la Psicólogo/a ha de estar profesionalmente preparado y especializado en la utilización de métodos, instrumentos, técnicas y procedimientos que adopte en su trabajo. Forma parte de su trabajo el esfuerzo continuado de actualización de su competencia profesional. Debe reconocer los límites de su competencia y las limitaciones de sus técnicas” (artículo 17).

En el artículo 18 se señala: “Sin perjuicio de la legítima diversidad de teorías, escuelas y métodos, el/la Psicólogo/a no utilizará medios o procedimientos que no se hallen suficientemente contrastados, dentro de los límites del conocimiento científico vigente. En el caso de investigaciones para poner a prueba técnicas o

instrumentos nuevos, todavía no contrastados, lo hará saber así a sus clientes antes de su utilización”.

Por lo tanto, la necesidad de la actualización continua de los conocimientos y la necesidad de aplicar los tratamientos que se han contrastado de forma válida y con pruebas científicas que avalen su efecto son aspectos fundamentales a tener en cuenta en la práctica profesional del psicólogo o psicóloga.

Desde el ámbito internacional destaca el Código Ético de la American Psychological Association, APA (2002) que afirma que los psicólogos deben mantener y desarrollar su competencia (artículo 2.03) y su trabajo debe estar basado en el conocimiento profesional y científico establecido de la disciplina (artículo 2.04). El primer código ético de la APA se elaboró en 1953 y posteriormente fue modificado y se convirtió en las primeras pautas éticas que guiarían el ejercicio de la profesión (American Psychological Association, 1953).

En segundo lugar, los investigadores que llevan a cabo estudios empíricos deben someterse a unas pautas de integridad científica que son la base primera y fundamental del proceso del diseño de investigación. Si la planificación del proceso de investigación no tiene como raíz principal la ética y la integridad del investigador poco se podrá hacer después durante el proceso del diseño de investigación y, por supuesto, los resultados y su interpretación quedarán contaminados.

A veces, la falta de integridad es por desconocimiento de las implicaciones metodológicas que su conducta tendrá sobre el resultado final al procesar y analizar los datos (conocido como ‘grados de libertad del investigador’) y otras veces se debe directamente a una conducta maliciosa e intencional dirigida a crear y/o falsificar los datos y obtener unos resultados concretos independientemente de la realidad del fenómeno (se trata del fraude científico).

Por ejemplo, desde el punto de vista de la formación del investigador o investigadora y también del lector o lectora activo, desconocer las implicaciones que tendrá sobre los resultados la baja potencia estadística de la prueba estadística también es una cuestión que afecta a la ética del investigador o investigadora y a los lectores y lectoras. El investigador / la investigadora es el responsable de su formación así como responsable de poner todos los elementos metodológicos que aseguren la calidad del proceso de diseño de investigación del estudio llevado a

cabo. Como lector o lectora de literatura científica, el profesional debe tener la suficiente formación en metodología que asegure que puede llevar a cabo una lectura de los informes científicos más allá de una comprensión sustantiva; es decir, es necesario que sepa valorar los hallazgos de una forma crítica o activa que le permita emitir un juicio sobre la validez de los resultados y sobre las afirmaciones que se mencionan en los informes. El saber científico requiere una constante actualización y requiere las competencias necesarias para elaborar un análisis de la anatomía metodológica del proceso del diseño de investigación. No todo lo que se publica tiene la calidad suficiente para aportar pruebas o evidencia de un fenómeno y, además, la calidad de los estudios no es del todo o nada sino que se puede jerarquizar y exige que los profesionales sepan valorar el grado de confianza que puede tener la información que aporta un artículo o un informe de investigación.

Grados de libertad del investigador o investigadora

La creciente sofisticación estadística, metodológica y técnica facilita que ahora sea el momento de abrir (de nuevo) el debate sobre la calidad de los resultados de la Ciencia, aportando pruebas o evidencia empírica sobre los elementos que podrían afectar a las decisiones del investigador a lo largo del proceso de diseño de investigación y que se denomina como “grados de libertad del investigador o investigadora”. No se trata de fraude científico (en el sentido de intencionalidad en la conducta para crear resultados falsos o para falsificarlos) pero sí de un tema vinculado con la ética y la integridad del investigador. Se trata de decisiones que puede tomar el investigador basadas en la flexibilidad que tiene a lo largo del proceso del diseño de investigación y que trae de nuevo a la luz la importancia de la ética, la integridad y los valores morales en la Ciencia; tres piezas clave para hacer buena Ciencia.

Por lo tanto, los denominados “grados de libertad del investigador o de la investigadora” son decisiones que toma el investigador no dirigidas intencionadamente al fraude o a falsificar los datos, pero sí que podrían ir dirigidas hacia la búsqueda del resultado estadísticamente (conocido como “*p*-hacking”, se podría traducir como la piratería del valor *p*) porque así se garantiza en gran medida la publicación del trabajo (provocando a su vez el sesgo de publicación: se publican mucho más los estudios con resultados estadísticamente significativos),

pensando (justificándose el propio investigador) que se trata de una mejora del diseño del estudio, pero tomando las decisiones a posteriori (una vez ya se dispone de resultados y por lo tanto si debe ajustarlos para lograr que el hallazgo sea estadísticamente significativo). La conducta de *p*-hacking supone que las decisiones están determinadas única y exclusivamente por el valor *p* obtenido en el estudio y se dirige todo el esfuerzo o conducta hacia lograr obtener un valor de *p* estadísticamente significativo ($p < .05$).

Con la conducta del *p*-hacking los investigadores no son conscientes de que sus decisiones a lo largo del proceso del diseño de investigación podrían conducir a lo que se llama un “resultado falso positivo” (informar de un resultado estadísticamente significativo cuando en realidad es falso).

Pero, la conducta de *p*-hacking en sí misma no es el problema que tiene la Ciencia sino que es un síntoma de la ‘cultura de la publicación’ (se publican en mayor medida los estudios con resultados positivos o estudios donde se detectan efectos estadísticamente significativos) y, también, es un síntoma de la ‘cultura de la investigación’ (forma de planificar y hacer diseño de investigación). Por ejemplo, los diseños de los estudios suelen caracterizarse por:

- Estudios con muestras pequeñas (problemas con la validez de conclusión estadística)
- Escasa potencia estadística (problemas con la validez de conclusión estadística)
- En muchas ocasiones tamaños del efecto pequeños (problemas con la validez de conclusión estadística)
- Sesgo de selección (problemas de validez externa)
- Falta de planificación de los elementos fundamentales que forman parte del diseño y el análisis estadístico: alfa, beta, efecto estimado, potencia deseada, *N* o número de observaciones (muestra) a recoger (problemas de validez)
- Desconocimiento de la magnitud del efecto que se desea detectar (efecto estimado) y de ahí su falta de consideración en la planificación del estudio y en la estimación de la potencia estadística a priori (problemas con la validez de conclusión estadística)
- Instrumentos de medida con escasa fiabilidad o consistencia interna (problemas con la validez de constructo)
- Falta de información sobre la elaboración, ejecución del estudio y redacción de los resultados (problemas en el informe o reporte de los hallazgos de investigación)

- Mala comprensión de lo que representa un resultado nulo o resultado negativo (mantener la hipótesis nula en el contraste estadístico) y lo que significa un resultado estadísticamente significativo o resultado positivo (rechazar la hipótesis nula en el contraste estadístico) dentro del procedimiento clásico de significación de la hipótesis nula (NHST) (De ahí las interpretaciones incorrectas que se realizan de los resultados obtenidos con las pruebas de contraste estadístico problemas de comprensión y educación estadística).

No todos los grados de libertad del investigador son prácticas de investigación cuestionables, pero sí es cierto que en la toma de decisiones del investigador a lo largo del proceso del diseño de investigación se puede optar por una decisión cuestionable desde el punto de vista ético, ya que distorsiona la realidad buscando los resultados estadísticamente significativos. Y estas conductas también afectan a la probabilidad de replicar dichos hallazgos. Es cierto que hay muchas razones basadas en el rigor metodológico que permiten eliminar una hipótesis una vez ejecutado el estudio, pues por ejemplo, si la consistencia interna de un instrumento no fue la adecuada entonces se puede argumentar por qué esa variable no fue estudiada o si, finalmente no se pudo completar la recogida de muestra planificada en un grupo se puede argumentar por qué las puntuaciones de ese grupo no se tuvieron en cuenta en el análisis de los datos. Pero todas esas conductas deben justificarse en la redacción del propio estudio desde un punto de vista metodológico y deben basarse en criterios fijados priori (es decir, antes de ejecutar el estudio) que el investigador deberá dejar reflejado en el denominado protocolo de su estudio (elaborado antes de realizar la investigación) y nunca se deben realizar maniobras o ajustes forzados de los datos (por ejemplo, eliminando o añadiendo datos sin dejar constancia de argumentos que lo justifiquen) para poder concluir que se ha comprobado la hipótesis de investigación, ya que se rechazó la hipótesis nula.

En resumen, el investigador cuando lleva a cabo un estudio quiere confirmar sus hipótesis y piensa que a lo largo del proceso del diseño de su investigación puede tomar ciertas decisiones para mejorar dicho diseño, ya que con ello aportará resultados válidos y, además, aumenta la probabilidad de que su trabajo sea publicado (Smaldino y McElreath, 2016). En cualquier caso, que haya o no voluntariedad en las malas prácticas de investigación lo importante es que se dan y pervierte el alcance y la validez del conocimiento científico (Frías-Navarro, Pascual-

Llobell, Pascual-Soler, Perez-Gonzalez, & Berrios-Riquelme, 2020). Ante esta situación, la denominada 'meta-investigación' (investigación sobre la investigación ya realizada) trata de describir la realidad de las actuaciones que se llevan a cabo en la Ciencia para aportar herramientas de educación y re-educación metodológica que garanticen la ejecución de investigaciones fiables y válidas (Ioannidis, 2018; Ioannidis, Fanelli, Dunne, y Goodman, 2015). Otra cuestión también importante para reflexionar es cómo valorar el trabajo de los científicos que hacen meta-investigación (es decir, ¿quién controla al controlador?).

Por lo tanto, elaborar una investigación exige una conducta íntegra y ética del investigador que comienza con la preparación metodológica adecuada que garantice la calidad de los hallazgos y su adecuada interpretación. Y, leer la literatura científica exige acercarse a su contenido desde una perspectiva crítica o activa que requiere también la formación metodológica adecuada que garantice la posibilidad de detectar sesgos o problemas que amenazan a la validez de los resultados presentados en el artículo o informe de investigación.

Por todo ello, la transparencia del informe de investigación en todos sus apartados (también aportando la base de datos del estudio en repositorios o plataformas especializadas en abierto o en anexos *on-line* mantenidos por las propias revistas donde se publica el artículo), el pre-registro del estudio (elaborar un protocolo y publicarlo) y la publicación del informe final en abierto en los repositorios (donde se somete de nuevo y de forma perpetua a la evaluación de los investigadores y lectores) son fundamentales para evitar el sesgo en la conducta del investigador y para posibilitar los estudios de meta-ciencia.

También los estudios de meta-análisis mejorarían con esas prácticas de transparencia, pues muchas veces se debe recurrir a estimaciones sesgadas de los estadísticos, ya que el artículo no detalla los datos necesarios para hacer un cómputo directo del tamaño del efecto. Por ello, es necesario informar siempre de las medias junto con su desviación típica y tamaño muestral e informar de los resultados de correlación junto con su tamaño muestral. De este modo se facilita realizar estimaciones más precisas del tamaño del efecto así como estimar su intervalo de confianza, mejorando la interpretación sustantiva de los hallazgos.

La perspectiva de la Ciencia en Abierto ('Open Science') es una de las herramientas que permite el cambio en la conducta del investigador e Internet es la vía que la hace posible, ya que proporciona una infraestructura que antes no había y facilita que la comunidad científica acceda a todos los elementos del proceso del diseño de investigación y con ello a su valoración. Sin embargo, no solo con la Ciencia en Abierto se puede acabar con la cultura de la publicación y la cultura de la investigación. La educación y la re-educación estadística entre el alumnado, investigadores y profesionales para que lleven a cabo una reflexión crítica sobre los elementos del diseño que afectan al resultado final del proceso de diseño de investigación son fundamentales y requisitos imprescindibles.

La integridad científica también se aprende con la educación y la re-educación estadística, ya que se visualiza y se comprende el alcance de las decisiones que se toman a lo largo del proceso de diseño de un estudio y se reconoce (se comprende y se valora) cómo las conductas del investigador deben estar siempre justificadas metodológicamente y, sobre todo, se comprende la debilidad del ser humano que le puede llevar a tomar decisiones (conscientes o inconscientes) que quedan fuera de la integridad científica.

Método científico

La Psicología como ciencia empírica adquiere su conocimiento sobre los constructos que le son propios, planteando hipótesis y aplicando una serie de técnicas para controlar, recoger y analizar los datos que le permitirán contrastar dichas hipótesis con la realidad del fenómeno.

Las tareas que el método científico implica se manifiestan en diversas modalidades¹ de investigación o métodos dependiendo del planteamiento teórico de la investigación que guiará el tipo de hipótesis formuladas y las técnicas para recoger los datos. En la elaboración de este material se utiliza el término de metodología para abordar las diferentes modalidades de expresión del método científico.

¹ En general, el término *método* se utiliza indistintamente para referirse al método científico y a la modalidad que adopta el método para estudiar el problema concreto objeto de investigación. Para una revisión del tema consultar [Mayor y Pérez \(1989\)](#) quienes optan por mantener el término 'método' para denominar a la globalidad del método científico-positivo y para sus grandes alternativas: método experimental *versus* métodos no experimentales.

La Ciencia es un conocimiento organizado basado en la observación sistemática. Su meta es aplicar el método científico en la búsqueda de la ‘mejor’ solución para los problemas de investigación y también, en la búsqueda de la ‘mejor’ evidencia o pruebas científicas disponible hasta el momento.

El denominado método científico (del griego *-meta*, hacia, a lo largo y de *-odos*, camino y del latín *scientia*: conocimiento; ‘camino hacia el conocimiento’) supone buscar un hallazgo mediante la observación sistemática siguiendo una serie de etapas ordenadas que forman el proceso de la investigación. Desde el punto de vista de la Ciencia² se trata de un procedimiento que abarca el ciclo completo de la investigación de un determinado problema, avanzando hacia el conocimiento de forma sistemática, racional, rigurosa y crítica.

Muy brevemente, el ciclo del método científico incluye formular un problema de investigación de forma precisa planteando con ello una ‘necesidad de conocimiento’ que desemboca en la ‘formulación de una hipótesis’ que habrá que ‘contrastarla con la realidad’ que se observa del fenómeno en el estudio, buscando evidencias o pruebas que ayudarán al científico o científica a determinar el valor de las hipótesis y del conocimiento producido con la investigación. Los hallazgos obtenidos tras aplicar el proceso del método científico (proceso del diseño de investigación) serán la base de la formulación de un nuevo problema o de una nueva idea (nueva necesidad de conocimiento) que será también sometido al estudio científico con las pautas de dicho método (ver Figura 1). De este modo se produce el proceso cíclico de la acumulación del conocimiento científico y el progreso de la Ciencia.

² Ciencia en griego “episteme”, en latín “scientia”, en anglo-francés “science”, en italiano “scienza” y en alemán: “wissenschaft”.

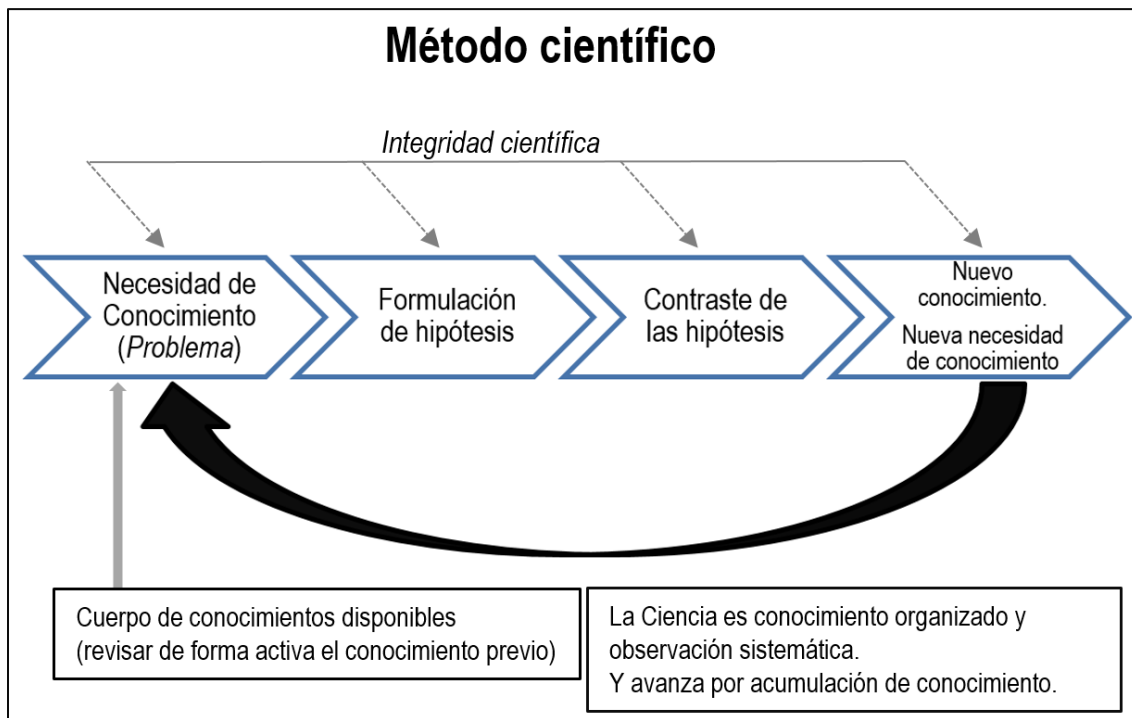


Figura 1. Método científico

Una investigación científica, por lo tanto, se caracteriza por ser *generalizable* (los resultados se pueden aplicar a otras muestras de sujetos, a otros momentos o a otros contextos), *empíricamente replicable* (se pueden replicar los efectos de los experimentos), *transparente* (se detalla lo que se hizo y por qué se hizo), por apoyarse en los resultados de la *investigación anterior* (conocimiento previo) y por *generar nuevas ideas* para aumentar el conocimiento de un fenómeno y avanzar en el conocimiento científico de los fenómenos.

No todos los profesionales de la Psicología o de los diferentes ámbitos de actuación van a realizar experimentos o investigaciones en su trabajo, pero sí todos los profesionales deben buscar la mejor evidencia científica disponible dentro de su área de trabajo (lecturas activas de artículos, informes, base de datos, protocolos, resultados...). Y para decidir qué evidencia científica es la de mayor calidad científica necesariamente necesitan conocimientos de metodología de investigación y del proceso del diseño de investigación. La lectura de los textos científicos (artículos, informes ...) se ha convertido en una tarea de 'valoración crítica' o activa que utiliza criterios objetivos de calidad de los hallazgos que están vinculados con todos los elementos del proceso de investigación que implica un diseño.

Hay muchas clases de investigación que aportan conocimiento a las Ciencias Sociales y de la Salud y no todas implican el uso de la estadística, pero también es cierto que la mayoría de las revistas incluyen un gran número de artículos con trabajos empíricos que hacen uso del diseño de investigación y la estadística.

La calidad del conocimiento científico generado en una disciplina requiere que los investigadores planifiquen adecuadamente su investigación, la ejecuten eficientemente, analicen los datos correctamente, interpreten bien los resultados y presenten de forma clara las conclusiones. Por supuesto, la calidad del conocimiento científico ya elaborado o publicado también depende de las lecturas y valoraciones de los científicos y profesionales, ya que, entre otras funciones, son evaluadores de la validez y credibilidad de los hallazgos publicados y ante hechos sospechosos de fraude o contaminación de los resultados deben informar a la comunidad científica para que se determina si un artículo o trabajo es retirado de la revista y justificar el porqué de dicha decisión. La literatura publicada o accesible a los lectores estará sometida a evaluación de forma perpetua por los miembros de la comunidad científica.

El desarrollo de la denominada 'Práctica Basada en la Evidencia' exige que el investigador, el profesional o el consumidor de literatura valoren los resultados de forma crítica y sean conscientes de que toda la información no vale lo mismo, es decir, sus resultados no tienen la misma calidad o validez (ver Figura 2) (Frías-Navarro y Pascual-Llobell, 2003; Pascual-Llobell, Frías-Navarro y Monterde-i-Bort, 2004). Es necesario "separar el grano de la paja", es decir, valorar la calidad de los hallazgos para poder deshechar los estudios cuyos resultados tienen problemas de validez (presentan sesgos) y atender de forma detenida a los hallazgos de las investigaciones que pasan el filtro de la lectura crítica del profesional o investigador/investigadora.

La validez de los resultados de los trabajos se puede jerarquizar en función del control del sesgo que se haya realizado en el diseño del estudio. Y es importante tener en cuenta que la opinión de la autoridad ('opiniones de los expertos') basada en la propia experiencia sin aportar pruebas o evidencia científica con calidad ha dejado de ser una fuente de información con garantías de veracidad.

- El movimiento de la denominada “Práctica Basada en la Evidencia” (PBE) exige que los investigadores lleven a cabo el proceso del diseño de la investigación (método científico) maximizando el control de los sesgos, garantizando la calidad o validez de los resultados y el avance del conocimiento científico.
- Los resultados de las investigación no tienen todos la misma calidad. La calidad metodológica del diseño de la investigación se puede jerarquizar valorando los elementos que intervienen durante todo el proceso del diseño de la investigación.

SIEMPRE ES NECESARIO APLICAR EL MÉTODO CIENTÍFICO EN LA BÚSQUEDA DE LA MEJOR EVIDENCIA CIENTÍFICA

- La calidad del conocimiento científico generado en una disciplina requiere que los investigadores:
 - ✓ Planifiquen adecuadamente su investigación.
 - ✓ Ejecuten la investigación eficientemente.
 - ✓ Analicen los datos de forma correcta.
 - ✓ Interpreten bien los resultados.
 - ✓ Presenten de forma precisa y clara las conclusiones en el informe.
 - ✓ Y que el contenido del informe sea transparente y
 - ✓ que todas las decisiones adoptadas por el investigador queden registradas en el informe o sus anexos y puedan ser accesibles para posibilitar su lectura crítica o activa.

Figura 2. Método científico y Práctica Basada en la Evidencia

Por lo tanto, las decisiones de los profesionales exigen cambiar el uso de las opiniones de los expertos como opiniones clave por el uso de opiniones basadas en el método científico y por saber valorar la calidad de las evidencias o pruebas halladas para abordar los problemas y obtener las conclusiones sobre intervención, diagnóstico, etiología o pronóstico de un fenómeno. Conviene, por ello, ser muy cauteloso cuando se valoran terapias sin pruebas de eficacia o efectividad realizadas con el método científico, ya que, si no se demuestra la validez científica de sus resultados, deberán ser clasificadas como pseudoterapias y consideradas como hechos no probados, sin el sello de calidad que otorga la ejecución escrupulosa de todos los pasos del proceso del método científico (proceso del diseño de investigación).

Como conclusión, destacar, en primer lugar, que comprender cómo se ha construido la investigación científica exige conocer en profundidad los elementos que determinan el método o metodología de investigación y el alcance de las

interpretaciones causales o no de los hallazgos. Conocer los fundamentos de la metodología de investigación es un requisito para poder producir investigaciones con resultados válidos y acumular conocimiento científico. Y, también, es un requisito para realizar lecturas críticas de los informes de investigación.

En segundo lugar, hay que tener siempre en cuenta que la calidad de la producción científica no siempre cumple los criterios de validez (Imai, King y Stuart, 2008; Onwuegbuzie, 2001). Los problemas de comprensión de la herramienta estadística y del diseño de investigación son una de las principales causas que invalidan los resultados de la investigación (Belia, Fidler, Williams y Cumming, 2005; delMas y Liu, 2005; Falk, 1986). Productores y consumidores de la información científica necesitan la educación estadística y necesitan conocer y aplicar los elementos que la nueva reforma estadística ha introducido desde la década de los noventa del siglo XX (Hancock y Mueller, 2010; Huck, 2007).

Y, en tercer lugar, comprender el proceso de contraste de hipótesis estadísticas es fundamental para el investigador o investigadora, que por supuesto lo debe conocer y saber aplicar, y también para el lector o lectora de informes y artículos de naturaleza empírica, ya que es el procedimiento de análisis estadístico más utilizado descrito en dichos informes. Saber interpretar los valores p de probabilidad, el tamaño del efecto y sus intervalos de confianza son competencias básicas del profesional de aquellas disciplinas donde se aplica la inferencia estadística tradicional y la estimación de efectos.

Reforma estadística y Práctica Basada en la Evidencia

Conocer el tamaño del efecto y la precisión de su estimación puntual con el intervalo de confianza permite realizar interpretaciones sobre la magnitud de las diferencias detectadas y no sólo sobre el grado de significación estadística. Las políticas editoriales de prácticamente todas las revistas científicas se han sumado a la denominada 'reforma estadística' modificando las instrucciones a los autores para el envío de sus manuscritos.

El movimiento de la Medicina Basada en la Evidencia (MBE) y, en general, el de la Práctica Basada en la Evidencia (PBE) tiene un punto de arranque en la toma de conciencia de los problemas vinculados al procedimiento de significación de la

hipótesis nula como único medio para producir descubrimientos, enfatizando el uso de las revisiones sistemáticas y la estimación del tamaño del efecto (Sackett, Strauss, Richardson, Rosenberg y Haynes, 2000). Además, el auge del movimiento de la Práctica Basada en la Evidencia (Frías-Navarro y Pascual-Llobell, 2003) centra la atención en obtener las mejores pruebas o evidencia facilitando así la interpretación de los hallazgos dentro del contexto específico de efectos al que pertenece la hipótesis de investigación. Es decir, un tamaño del efecto pequeño en términos estadísticos podría ser grande en un sentido sustantivo, del mismo modo que un tamaño del efecto grande en términos estadísticos podría ser pequeño en el sentido sustantivo, clínico o de utilidad para el profesional o investigador.

La reforma estadística cambia el punto de mira desde “cómo es de probable o improbable el resultado muestral” hacia dos cuestiones principalmente: “cómo es de grande el tamaño del efecto detectado” y “si se puede replicar”. Es decir, hay que ‘evaluar’ el valor del tamaño del efecto estimado y su utilidad (su grado de importancia práctica, clínica o sustantiva) y para ello es necesario considerar el contexto de la investigación y comparar de forma explícita y directa los resultados de un estudio con los obtenidos en el área de investigación donde se enmarca el trabajo, es decir, se trata de contextualizar los efectos (Cumming y Finch, 2005; Frías-Navarro, Pascual-Llobell, y García-Perez, 2000; Wilkinson & the Task Force on Statistical Inference, 1999). Además, la replicabilidad del efecto supone evaluar cómo de estables son los efectos en la literatura revisada y por lo tanto evaluar en qué medida son efectos directamente comparables (Pascual-Llobell, García-Perez, y Frías-Navarro, 2000).

En definitiva, el nuevo comportamiento del investigador supone desarrollar el denominado “pensamiento meta-analítico” (Cumming y Finch, 2001). En este sentido, los investigadores son quienes deben planificar sus hipótesis en función de los efectos previos detectados en la literatura o en el conocimiento teórico e interpretar sus hallazgos dentro de dicho contexto de efectos. El ‘pensamiento dicotómico’ (vinculado a la comprobación tradicional de la hipótesis nula: mantenerla o rechazarla) es reemplazado por un ‘pensamiento meta-analítico’ (valoración de la magnitud del tamaño del efecto y su intervalo de confianza y contextualizar los efectos) que facilita llegar a una interpretación sustantiva de los hallazgos integrada en el contexto de efectos que hasta el momento se han

obtenido en las investigaciones. Dicha interpretación sustantiva deberá ser valorada como útil (grado de utilidad) por el profesional desde su juicio clínico.

La elaboración de estudios de meta-análisis y sobre todo el uso de sus resultados para planificar el tamaño del efecto de un estudio primario y contextualizar los efectos adquieren una especial relevancia entre las indicaciones de la reforma estadística (Sánchez Meca y Botella, 2010). El cambio implica que los investigadores planteen su diseño estadístico en términos de estimación de efectos y no solo en lograr la significación estadística. Y este cambio no está siendo fácil para los investigadores y las investigadoras tal y como demuestran los estudios sobre el uso del tamaño del efecto y sus intervalos de confianza en las publicaciones científicas (Frías-Navarro, 2011).

La edición del Manual de la *American Psychological Association* de 2010 mantiene su énfasis en la denominada reforma estadística, destacando el uso de los tamaños del efecto y sus intervalos de confianza y las técnicas bayesianas tratando de minimizar la confianza excesiva que los investigadores tienen sobre las pruebas de significación estadística y las decisiones dicotómicas apoyadas en los valores p de probabilidad (Hoekstra, Finch, Kiers, y Johnson, 2006). El grupo de trabajo de inferencia estadística de la *American Psychological Association* (Wilkinson & APA Task Force on Statistical Inference, 1999) y las tres últimas ediciones del Manual de la APA (*American Psychological Association*, 2001, 2010, 2019) señalan de forma destacada que los investigadores deben estimar el valor del tamaño del efecto y su intervalo de confianza junto al valor p de probabilidad, siempre que sea posible. En definitiva, la *American Psychological Association* (APA, 2001) recomienda a los investigadores estimar y tomar en consideración el tamaño del efecto en el ámbito teórico y en el aplicado. Afirma el Manual que una característica esencial de la buena investigación es la de interpretar los tamaños del efecto en relación con los efectos previamente estimados e informados por la investigación que ya se ha realizado dentro del campo de estudio (pensamiento meta-analítico).

Las nuevas recomendaciones formuladas en la reforma estadística no llegan a consolidarse entre la conducta de los investigadores (Vacha-Haase, Nilsson, Reetz, Lance, y Thompson, 2000). Muy probablemente los investigadores aún no han

reflexionado sobre la cuestión de que “seguramente Dios ama al .06 (nivel de significación estadística) tanto como al .05” (Rosnow y Rosenthal, 1989, p.1277). Además, los programas estadísticos, como el SPSS, aún no han incorporado la estimación directa de los nuevos estadísticos como la *d* de Cohen o los intervalos de confianza de forma generalizada en todas las estimaciones puntuales de los estadísticos (por ejemplo, cuando se estima un coeficiente de correlación), dificultando con ello el cambio de la práctica estadística. Otros programas como Epidat (Santiago, Hervada, Naveira, Silva, Fariñas y cols., 2010), JASP, JAMOVI ya han incorporado la estimación del tamaño del efecto y, en gran parte, la estimación de los intervalos de confianza.

La reforma estadística también destaca la capacidad de leer la literatura científica de una forma crítica como una competencia esencial del consumidor o consumidora de informes de investigación que exige tener conocimientos de metodología para valorar los trabajos con investigación empírica. La literatura es el repositorio escrito del conocimiento. Su revisión ayuda a perfilar de manera más elaborada la hipótesis de investigación. Una buena hipótesis debe ser factible o abordable, interesante, novedosa, ética y relevante para generar conocimiento científico.

El propósito de la lectura crítica o activa es doble:

1. por una parte, para comprobar si los métodos de investigación aplicados producen información útil y, por otro lado,
2. para analizar si las conclusiones de los trabajos de investigación realmente pueden ser obtenidas a partir de los hallazgos obtenidos. Es decir, se trata de valorar la calidad o validez de los hallazgos tal y como han sido obtenidos con el diseño de investigación aplicado en el estudio que se lee.

Diseño de investigación

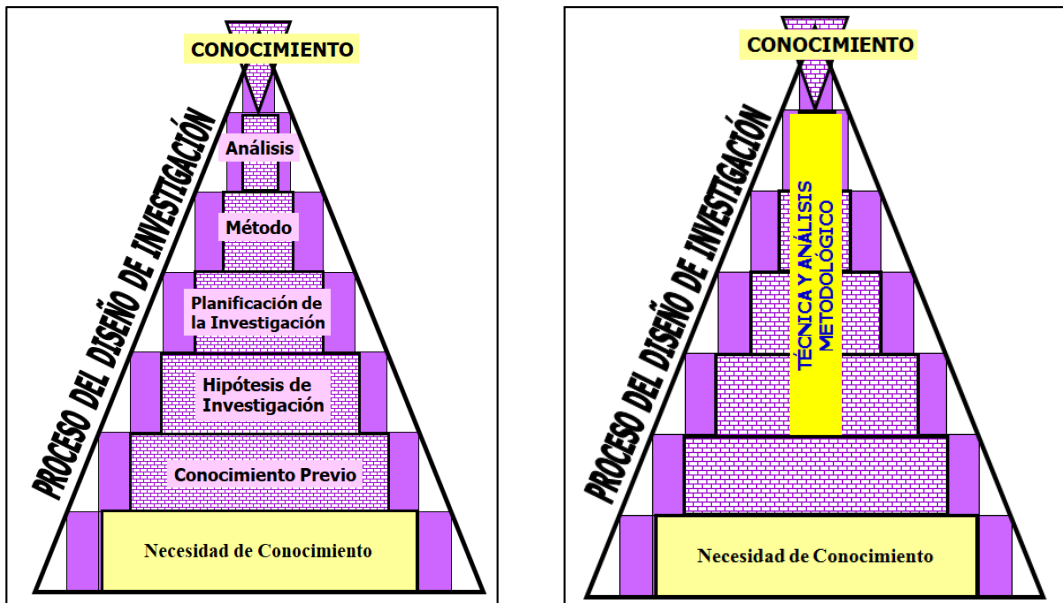
Dentro del ámbito metodológico cuando se habla de diseño de investigación existe cierto consenso en caracterizarlo como un conjunto de actividades dirigidas a resolver un problema concreto (*necesidad de conocimiento*), incluyendo como elementos propios del diseño desde el planteamiento teórico del problema de

investigación a partir del conocimiento previo y formulación conceptual de las hipótesis hasta su análisis estadístico, interpretación y discusión de resultados (ver Figura 3) (Arnau, 1981; Ato, 1991; Frías-Navarro, 2011; García, Frías-Navarro, y Pascual-Llobell, 2006; Kirk, 1995; Maxwell y Delaney, 2004; Maxwell, Delaney, y Kelley, 2018; Pascual-Llobell, Frías-Navarro, y García, 1996).

En definitiva, partiendo del concepto de Psicología como una actividad científica, cuyo campo de estudio es el comportamiento humano, y del método de investigación como uno de los caminos válidos para contrastar los enunciados científicos con la realidad, el término “diseño de investigación” implica la planificación de todos los elementos necesarios para contrastar de forma correcta (sin sesgo) la o las hipótesis del estudio con dicho método o alcanzar los objetivos propuestos en el estudio.

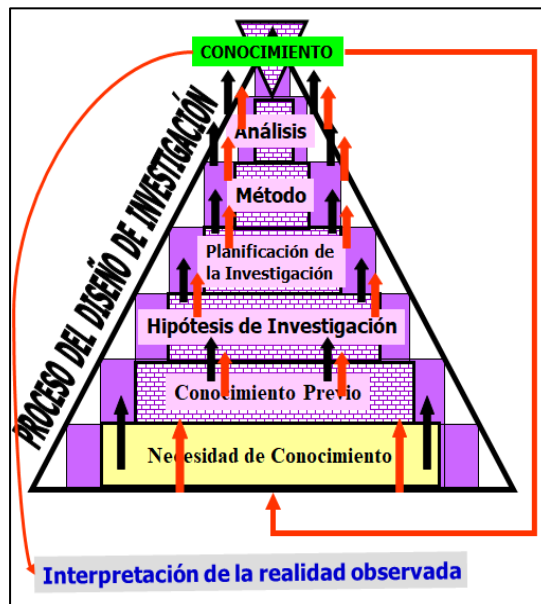
Desde el ámbito de la investigación empírica cuantitativa, la formulación de las hipótesis teóricas es, por supuesto, inherente al proceso de investigación dando sentido a la ejecución del experimento o estudio como un procedimiento de contrastación de hipótesis estadísticas y comparación de modelos explicativos, cuyo fin es poder determinar si las hipótesis teóricas se confirman o no (siempre con un riesgo de error en la decisión estadística). Es decir, el fin último del contraste de hipótesis estadísticas es aportar evidencia empírica sobre las relaciones teóricas entre las variables. Por lo tanto, trasladar una idea o un problema de investigación a un plan de trabajo requiere comprender los fundamentos del diseño de investigación y sus técnicas de análisis.

Por lo tanto, cuando el investigador plantea una ‘necesidad de conocimiento’ (problema de investigación) vinculada posteriormente a una hipótesis de investigación se inicia el proceso de diseño de investigación (ver Figura 3).



a

b



c

Figura 3. Pirámide del proceso del diseño de investigación

El proceso de diseño de investigación se inicia con la *necesidad de conocimiento* que tiene el investigador sobre un determinado constructo o problema psicológico (objetivo del estudio, por qué se lleva a cabo la investigación) y no termina hasta lograr cierto *conocimiento* válido sobre la realidad del fenómeno estudiado (ver Figura 3a).

Después de una etapa previa, donde se generan ideas sobre el fenómeno objeto de estudio, se concreta una “necesidad de conocimiento” vinculada al problema de investigación. Es decir, el investigador define lo que va a investigar y restringe el campo a una pregunta concreta, delimitando las variables y operacionalizando la pregunta que hay que resolver. Posteriormente se pone en marcha la técnica estadística y el análisis metodológico que requiera el problema de investigación planteado (ver Figura 3b).

Finalmente la valoración y la interpretación de los resultados obtenidos permiten elaborar el nuevo conocimiento, continuando de este modo el proceso cíclico de formación del saber científico (ver Figura 3c).

La pirámide invertida que aparece en las representaciones de la Figura 3 en la parte superior donde se ubica el ‘Conocimiento’ adquirido señala la dirección hacia una nueva pirámide (un nuevo estudio con su proceso de diseño de investigación) que será construida sobre la base del conocimiento previo obtenido con la investigación que se ha finalizado. De este modo, el saber científico se va acumulando con estudios que aplican el método científico y va formando el conocimiento científico sobre un determinado fenómeno o constructo.

Por lo tanto, la investigación científica implica un proceso de análisis de la realidad con el objetivo de dar respuesta a la necesidad de conocimiento planteada en una determinada investigación.

Conviene tener siempre muy presente que la formulación de las hipótesis teóricas, la recogida de los datos y el planteamiento de su análisis estadístico son los elementos que guían la elección del diseño de investigación más apropiado. Por lo tanto, es conveniente dedicar tiempo y esfuerzo a los pasos previos a la recogida de los datos (es decir, a la fase de planificación), ya que una vez recogidos y ejecutado el estudio poco se podrá hacer para mejorar el diseño del estudio.

La contrastación con la realidad supone poner en marcha técnicas de comprobación de hipótesis estadísticas que unidas al diseño seleccionado por el investigador o investigadora permiten alcanzar una explicación válida de los resultados obtenidos, comparando lo predicho por la teoría y lo manifestado en la realidad, valorando y controlando en el diseño aquellas variables que podrían confundir los resultados.

Como posteriormente se detallará, con la denominada metodología experimental y la cuasi-experimental el propósito es potenciar el efecto de aquella variable que se ha manipulado o *variable independiente de tratamiento* para estudiar su efecto en una determinada variable medida objeto de observación o *variable dependiente* y controlar la posible influencia de *variables extrañas* que no forman parte de las hipótesis y cuyo efecto sistemático (sesgo) sobre la variable dependiente confundiría la interpretación de los resultados. Cuando se trabaja con una metodología no experimental el propósito es potenciar la magnitud de la relación entre las variables que se van a estudiar, evitando por supuesto la presencia de variables contaminadoras o extrañas (también denominadas ‘terceras variables’ que están unidas al sesgo o a la contaminación de los datos).

Capítulo 2. Variables del estudio

Dolores Frías-Navarro

Universidad de Valencia

Citar el capítulo como:

Frías-Navarro, D. (2020). Variables del estudio. En D. Frías-Navarro y M. Pascual-Soler (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP>
Disponible en: osf.io/kngtp

Una variable es cualquier entidad que puede tomar diferentes valores. Es decir, una variable es una característica de un individuo o de un objeto que puede tener diferentes valores y por ello es lo opuesto a una constante. La conducta de un fenómeno suele estar provocada por multitud de factores o variables, definiéndose en el diseño de la investigación un reducido número de posibles variables que provocan su comportamiento, siendo el estudio de dichas variables el objetivo de la investigación.

Clasificación de las variables del estudio

Las variables del diseño de investigación pueden ser clasificadas por diferentes criterios, destacando el criterio metodológico que es uno de los más utilizados e inicialmente unido a los diseños con una metodología propiamente experimental (variable independiente, variable dependiente y variables controladas), el criterio manipulativo de la variable de intervención (variable independiente manipulada o variable independiente asignada) y el criterio estadístico (variables cualitativas / cuantitativas y variables discretas / variables continuas).

Criterio metodológico

Utilizando un *criterio metodológico*, vinculado a la metodología experimental, se distinguen tres tipos de variables en una investigación:

1. la o las variables independientes (VI): son las variables de tratamiento o los 'factores' del diseño de investigación. Se representan como A, B, C ...
2. la o las variables dependientes (VD): son las variables que se miden en el estudio (el resultado). Se representan en el diseño por $Y_1, Y_2 \dots$
3. la o las variables extrañas a los objetivos o hipótesis de investigación (VE): son las variables que se controlan en el diseño de investigación (variables controladas por ejemplo por la técnica de la aleatorización, la eliminación o la constancia) y también pueden ser controladas con el tipo de diseño elegido, ya que pueden actuar como factores del diseño o como fuentes de varianza cuyo objetivo es controlar el efecto de posibles variables extrañas. Se trataría de factores o variables que no son objeto de estudio en la hipótesis teórica planteada, ya que su función en el diseño es de control del efecto de 'terceras variables' (por ejemplo, el diseño de bloques o el diseño con variables covariadas) y no la de aportar una explicación teórica del fenómeno.

El papel metodológico que tiene cada tipo de variable en el diseño del estudio es asignado por el investigador o investigadora en función de sus intereses e hipótesis concretas. Así, la ansiedad puede ser una variable dependiente si se utiliza como variable medida o resultado, puede ser una variable independiente categorizada en diferentes condiciones (por ejemplo, ansiedad baja, ansiedad media y ansiedad alta) o podría ser una variable extraña que contaminaría los resultados si no se controla con alguna técnica (por ejemplo, se puede utilizar como factor de bloqueo o como una variable covariada de ajuste estadístico).

Conviene tener muy presente que dado que la Psicología trabaja con constructos o fenómenos no observables como la depresión, la ansiedad, la autoestima, la anorexia, el prejuicio, el racismo, la homofobia..., es necesario valorar las variables del estudio en su componente de constructo y en su

componente operacionalizado tal y como es tratado de forma empírica en la investigación. Por ejemplo, si se desea estudiar el nivel de ansiedad que produce la cercanía de una serpiente será necesario que en el estudio se operacionalice la medida del constructo de ansiedad, por ejemplo por la tasa cardíaca, y la cercanía al estímulo se puede medir en centímetros.

Variable independiente

Las variables independientes o factores del diseño de investigación (simbolizados por letras mayúsculas como factor A, factor B, factor C...) representan la intervención o tratamiento (o el grupo de comparación) o las fases de medición de una variable que se llevan a cabo si se trata de un diseño de investigación con medidas repetidas. El investigador / lector debe describir / detectar la variable constructo y la variable tal y como se ha operacionado en el diseño de esa investigación.

Cada variable independiente está formada por al menos:

- 1- Dos condiciones o grupos, en el caso del diseño de comparación de grupos más simple con una sola medición (conocido como '*diseño-entregupo*'). Las condiciones o grupos se simbolizan por las letras de los factores, pero en minúsculas, por ejemplo a_1 y a_2 , b_1 y b_2 , c_1 y c_2 ...
- 2- O puede tratarse de una variable independiente con al menos dos mediciones, en el caso de los '*diseños de medidas repetidas o diseños intra-sujetos*' más simples (a_1 y a_2), actuando cada medición como una condición del factor (es decir, como medición primera a_1 y medición segunda a_2).
- 3- El diseño también puede estar formado, al menos, por una variable o factor 'entre-sujetos' y por, al menos, una variable o factor de medidas repetidas o intra-sujetos, formando los denominados '*diseños mixtos o de medidas parcialmente repetidas*'. Este tipo de diseños tendrá unas fuentes de varianza o variabilidad entre-grupos y otras fuentes de varianza intra-grupos y un término de error diferente para cada una de esas fuentes de varianza.

En función del número de variables independientes o factores que tiene el diseño de investigación, los diseños se pueden clasificar en '*unifactoriales*', diseños

con una sola variable independiente, y diseños '*factoriales*', diseños con más de una variable independiente o factor.

Teniendo en cuenta el número de observaciones que se recogen en cada condición de la variable independiente, los diseños entre-sujetos se pueden clasificar, además, en '*diseños ortogonales*' cuando el número de observaciones es el mismo en los dos grupos ($n_1 = n_2$) y '*diseños no ortogonales*' ($n_1 \neq n_2$) cuando el número de observaciones es diferente. En los diseños intra-sujetos se recoge el mismo número de observaciones en la medición 1 y en la medición 2 para poder ejecutar los análisis de medidas repetidas.

Variable independiente: manipulada / asignada

Las variables independientes del experimento pueden ser clasificadas utilizando diferentes criterios como el criterio manipulativo que distingue entre:

- 1- Variables independientes '*activas*' (variables manipuladas), cuyos niveles o condiciones son provocadas por el investigador o investigadora, como por ejemplo el tipo de fármaco o el carácter de las instrucciones dadas a los sujetos.
- 2- Variables independientes '*asignadas*' (también conocidas como variable observada, variable seleccionada o variable de clasificación) cuyas condiciones reflejan diferentes aspectos de la variable que no son manipuladas por el investigador o investigadora en el estudio sino que son medidas o registradas tal y como han sucedido en el sujeto, como por ejemplo la edad, el sexo, la profesión o el nivel de glucosa en sangre.

A continuación se describen los dos tipos de variables independientes con más detalle.

Variables independientes activas. Las variables independientes pueden ser "activas" cuando son variables manipuladas, es decir, son variables cuyas condiciones son creadas deliberadamente por el investigador o investigadora (son variables cuyos efectos están bajo el control del investigador o investigadora que puede ponerlas o quitarlas del diseño) para ver sus efectos sobre las variables dependientes o pueden ser variables '*asignadas*' (el investigador o investigadora no controla sus efectos). Una variable independiente manipulada puede ser un

tratamiento farmacológico, una terapia psicológica, un programa de intervención social o una tarea experimental que se crea para observar cómo reacciona el sujeto ante un determinado estímulo. En definitiva, supone introducir una variable en la vida del sujeto para ver su efecto o resultado sobre sus respuestas o variable medida (variable dependiente).

Por lo tanto, los efectos de las variables independientes activas están bajo el control del investigador o investigadora que introduce deliberadamente diferentes condiciones de tratamiento para observar cómo afectan a la variable dependiente, es decir, a la variable medida. Esto permite la generación controlada de datos, ya que la variabilidad de la variable independiente de tratamiento está determinada por el investigador o investigadora que puede hacer variar los valores de la variable independiente o decidir qué valores adoptará e incluso decidir cuándo la introduce y cuándo la retira del estudio.

Variables independientes asignadas. Las variables independientes “asignadas” son variables no manipuladas directamente por el investigador o investigadora sino que sus niveles o condiciones son seleccionados por el investigador o investigadora para observar su relación con otras variables. Aquí el investigador o investigadora selecciona ciertos niveles de una variable según sus intereses de investigación, pero no puede manipularlos.

Cuando no existe una intervención activa (manipulación) de la variable independiente (metodología no experimental) por parte del investigador o investigadora sino que se trata de un rasgo o de una exposición a una condición que ya posee el sujeto (sin manipulación) y que es seleccionada por el investigador o investigadora como objetivo de estudio entonces la variable independiente se conoce como la variable *predictora* y la variable medida como la variable *predicha*. En definitiva, los niveles de la variable independiente asignada los selecciona el investigador o investigadora (por cuestiones teóricas o de interés teórico) del conjunto de niveles que ya posee la variable predictora o factor. Por ejemplo, el sexo, el nivel educativo, la edad, la profesión, el estatus económico, la estructura familiar... son variables predictoras asignadas que son seleccionadas por el investigador o investigadora debido a su interés teórico.

Variable dependiente

La variable dependiente es la variable medida objeto de observación en el diseño de investigación; son los resultados del estudio. Se representa con la letra Y. Es el resultado que se mide en el estudio. Es el desenlace del experimento.

Por ejemplo, la variable dependiente puede ser el grado de eficacia del tratamiento medido con el nivel de colesterol en la sangre, medido con una escala de ansiedad o medido con un cuestionario de auto-eficacia percibida después de recibir un determinado tratamiento o intervención. O quizás, puede ser el nivel de cortisol secretado que aparece en la orina después de someter al sujeto a una situación de estrés como puede ser un examen tal y como se desarrolla en el trabajo de García de la Banda, Martínez-Abascal, Riesco, y Pérez (2004).

El investigador o investigadora debe describir en el informe o artículo todas las variables constructo (independientes, dependientes y extrañas) y su operacionalización en el diseño de esa investigación. El lector o lectora debe detectar o reconocer las variables constructo y su operacionalización en el estudio para poder llevar a cabo una lectura activa o crítica y comprender los elementos implicados en el proceso de diseño.

En función del número de variables dependientes o variables medidas que tiene el diseño de investigación se pueden clasificar los diseños en diseños 'univariados' que son diseños con una sola variable dependiente utilizada en el análisis, y diseños 'multivariados', es decir, diseños con más de una variable dependiente medida que se utilizan de forma simultánea en el análisis de los datos. Los diseños unifactoriales y los factoriales pueden ser univariados o multivariados.

Variable extraña

Las variables extrañas a los objetivos de la investigación son aquellos otros factores que podrían tener algún impacto importante en los resultados del estudio. Es decir, es cualquier otra variable que también podría afectar a la variable dependiente medida o interactuar con la variable independiente, pero que no son objeto de estudio en la hipótesis de investigación.

Las variables extrañas incluyen:

- 1- las variables 'aleatorias' cuyo efecto no es sistemático y representan el error aleatorio que está presente en cualquier diseño de investigación y
- 2- las denominadas variables 'perturbadoras' (terceras variables) que deben ser controladas para potenciar la validez o calidad de los resultados.

Las "variables extrañas aleatorias" están siempre presentes en el diseño de investigación y es objetivo de la planificación del estudio minimizar el error aleatorio para crear grupos de comparación lo más equivalentes posibles en todas las variables excepto en aquella variable cuyo efecto se desea analizar en el estudio.

Las "variables perturbadoras" son las variables no controladas en el diseño de investigación y su presencia podría distorsionar la relación entre la variable independiente y la variable dependiente (ambas variables forman la hipótesis de investigación) y podrían provocar desequilibrios en los diferentes grupos o condiciones de intervención, creando grupos no equivalentes y dando lugar a un sesgo sistemático en los datos. En este caso, la validez interna del diseño estará claramente cuestionada.

Por lo tanto, la clave de las variables extrañas está en si el sesgo que provocan es aleatorio (error no sistemático que hay que minimizar en el diseño) o es error sistemático (error sistemático que siempre hay que controlar con alguna técnica).

Es necesario que esas variables perturbadoras sean controladas ("variables controladas") con el diseño de investigación ya sea con la técnica de la aleatorización (variables extrañas aleatorizadas), con la técnica de la eliminación de su efecto si es posible (mantener constante su efecto sobre todos los datos), con la técnica del control matemático de su efecto mediante el diseño (diseño con variables covariadas donde se realiza un ajuste matemático de los datos) o bien incluyéndolas en el modelo de investigación como un factor más en la ecuación estructural para controlar su efecto por constancia (variables extrañas controladas por bloqueo) tal y como sucede en el diseño de bloques.

Siguiendo el planteamiento de una tipología de las variables que forman parte del plan de investigación, Kish (1975) distingue dos clases de variables: las *variables explicativas o experimentales* y las *variables extrañas* (ver Figura 4).

- 1- Las “variables explicativas” o experimentales constituyen el objetivo del estudio y son la o las variables independientes (o variables predictoras, variables estímulo, variables de tratamiento) y la o las variables dependientes (variables pronosticadas o predichas, variables de respuesta, variables criterio) que forman la hipótesis de investigación. Como el propio autor señala, estas variables surgen del conocimiento y la comprensión del campo de estudio.
- 2- Además, señala el autor, existen otras tres clases de variables que son “extrañas a los objetivos de la investigación”: las variables controladas, las variables perturbadoras y las variables aleatorias. Se trata de otras fuentes extrañas de variación que el investigador o investigadora debe identificar para separar su variabilidad de la producida por las variables explicativas.

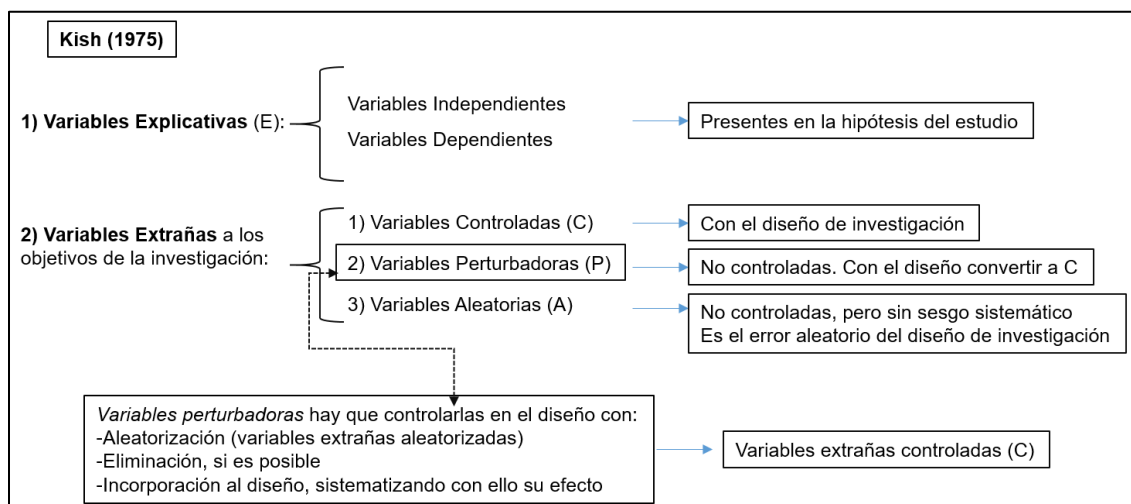


Figura 4. Variables de un diseño de investigación

Las ‘variables controladas’ son variables extrañas a los objetivos de la investigación que se controlan con el diseño de investigación con el objeto de disminuir los efectos del error aleatorio o eliminar los efectos de sesgo de las variables perturbadoras o para disminuir ambas situaciones.

Las ‘variables perturbadoras’, en cambio, son variables extrañas no controladas en el diseño de investigación y, por lo tanto, su efecto se puede confundir con las variables explicativas y provocar un efecto de confundido, interfiriendo la relación entre la variable independiente y la dependiente.

El objetivo de un eficiente diseño de investigación es controlar las variables perturbadoras, convirtiéndolas en variables controladas o, si es posible, convertirlas en variables aleatorizadas. Es importante reflexionar y detectar este tipo de variables (por ejemplo, tras efectuar la revisión del conocimiento previo) en la fase de planificación del diseño de investigación para que sea posible su control, tomando las decisiones antes de recoger los datos del estudio.

Por último, las ‘variables aleatorias’ son variables extrañas no controladas que se tratan en el diseño como errores del muestreo aleatorio (o de la asignación aleatoria del tratamiento) y se considera que no ejercen un sesgo sistemático sobre la relación entre la variable independiente y dependiente y el diseño de investigación debe minimizar la presencia de error aleatorio. Se trata del denominado error aleatorio que siempre está presente en la medida de la variable dependiente.

Por lo tanto, la variable independiente y la dependiente son las denominadas ‘variables explicativas’ del diseño de investigación y forman la hipótesis de investigación que describe el posible comportamiento del fenómeno investigado. Las variables controladas no se detallan en la hipótesis del estudio sino que se incorporan o se redactan cuando el investigador o investigadora describe el diseño de su investigación. Es muy importante leer (escribir) con detalle la hipótesis del estudio para poder detectar qué función tiene cada variable en el diseño de ese estudio.

La investigación puede plantear una “relación causal” entre las variables del estudio donde el resultado en la variable dependiente está causado por la variable independiente manipulada (metodología experimental) o bien puede tratarse de un estudio de “asociación” o de análisis de la magnitud de la relación entre las variables sin ningún tipo de interpretación causal entre las variables (metodología no experimental).

En resumen, las variables perturbadoras con efecto sistemático deben convertirse siempre en variables extrañas controladas dentro del diseño de investigación para garantizar la validez de los hallazgos. El sesgo sistemático provocado por el efecto de las variables extrañas no controladas es el enemigo número uno de la validez o calidad de los resultados. Si se conoce el efecto

sistemático de la o las variables extrañas y no se puede eliminar, emparejar o aleatorizar de forma directa entonces se debe incorporar al diseño como un factor más del estudio para controlar su presencia, ya que, en caso contrario, distorsionaría la verdadera relación entre las variables explicativas que forman la hipótesis de investigación (independiente y dependiente), confundiendo los datos obtenidos. Se trata de mantener constante el efecto de las variables extrañas en todos los grupos del diseño tal y como sucede en el diseño de bloques o tras aplicar la técnica de la aleatorización o residualizar estadísticamente la variabilidad que provocan utilizando en este caso diseños con variables covariadas.

Criterio estadístico

Por otra parte, el criterio estadístico clasifica a las variables por el tipo de unidad de medida y distingue entre *variables cualitativas* y *variables cuantitativas* y entre *variables discretas* y *variables continuas*.

Los valores de las “variables cualitativas” se encuentran en escala nominal y se diferencian por sus cualidades no por sus cantidades. No representan valores cuantitativos. Por ejemplo, el partido político, el estado civil, el sexo, la orientación sexual, la ciudad donde se vive, el color de los ojos, la profesión... Se les puede asignar un número, como por ejemplo 0 hombres y 1 mujeres para poder diferenciar las categorías, pero no tienen valor numérico. Si las variables nominales tienen dos categorías se les llama variables categóricas (por ejemplo el sexo de hombre y mujer) y si tienen más de dos categorías entonces son politómicas (por ejemplo el estado civil de soltero/a, casado/a, viudo/a).

Los valores de las “variables cuantitativas” son números y sus valores se pueden comparar entre sí como mayor o menor que otro valor. Por ejemplo, la edad, los ingresos económicos, el número de años recibidos de educación... Las variables cuantitativas pueden ser discretas y continuas.

Las “variables cuantitativas discretas” sólo pueden tomar valores enteros y tienen un número finito de valores entre dos valores dados. Por ejemplo, número de personas en el hogar, número de hijos, tamaño del municipio...

En las “variables cuantitativas continuas” el número de valores posibles entre dos valores dados es infinito (valores con infinitos decimales). Por ejemplo, la

altura, la longitud, el peso, la edad... En la práctica, las variables continuas que son el resultado de medir se redondean y se convierten en número finito de números enteros y las variables cuantitativas discretas son aquellas que tienen pocas categorías.

Las variables cuantitativas continuas pueden estar en escala de intervalo y en escala de razón. Las variables con escala ordinal son cuasi-cuantitativas, ya que sus valores son categorías como en las variables cualitativas nominales, pero sus valores pueden ser mayor o menor que otro valor como ocurre en las variables cuantitativas. Por ejemplo, la clase social entendida como baja (1) - media (2) - alta (3). El tipo de análisis que se puede realizar con los valores en escala ordinal es solamente de orden, pues no tiene sentido realizar operaciones aritméticas con los datos.

En las variables con escala de intervalo ya se dispone de una unidad de medida que es el intervalo entre dos valores y representa una distancia similar entre los valores en cada uno de los intervalos. El origen de la escala es arbitrario. Por ejemplo, ingresos económicos de 10000 euros, de 20000 o de 30000 euros. Es decir, se puede calcular la distancia o el intervalo entre cualquier par de valores de la variable. Permite realizar operaciones aritméticas como la suma o la resta.

En las escalas de razón existe una unidad de medida y un valor de cero absoluto (por ejemplo los ingresos realizados o los gastos, el peso, la edad, la estatura...). Con este tipo de escalas se admite cualquier tipo de operación aritmética.

Capítulo 3. Proceso del diseño de investigación

Dolores Frías-Navarro
Universidad de Valencia

Citar el capítulo como:

Frías-Navarro, D. (2020). Proceso del diseño de investigación. En D. Frías-Navarro y M. Pascual-Soler (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España.

DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> Disponible en: osf.io/kngtp

El proceso del diseño de investigación representa los pasos implicados en el método científico. A continuación se detallarán los elementos más importantes en cada una de las fases de dicho proceso o método científico.

Necesidad de conocimiento

La necesidad de conocimiento implica definir de forma clara, precisa y concreta el problema de investigación que se debe resolver, permitiendo así iniciar el proceso de búsqueda y localización de la información que facilitará la respuesta más adecuada, precisa y actual del problema planteado.

El problema de investigación (*necesidad de conocimiento*) debe expresar una relación (causal o no causal) entre dos o más variables y su planteamiento será claro y sin ambigüedades y debe permitir su verificación empírica. Además, el problema de investigación debe ser relevante, justificando el esfuerzo y la inversión económica y de trabajo humano que se invierte en su resolución.

El objetivo del estudio puede ser eliminar cierta incertidumbre que exista sobre el conocimiento de cierta temática o puede ser modificar y/o añadir nuevos conocimientos al área concreta de investigación (ver Figura 5). Además, es importante enmarcar el problema dentro de una teoría teniendo en cuenta que la Ciencia busca generalizar los hallazgos (que además sean replicables) y no se construye con hechos aislados.

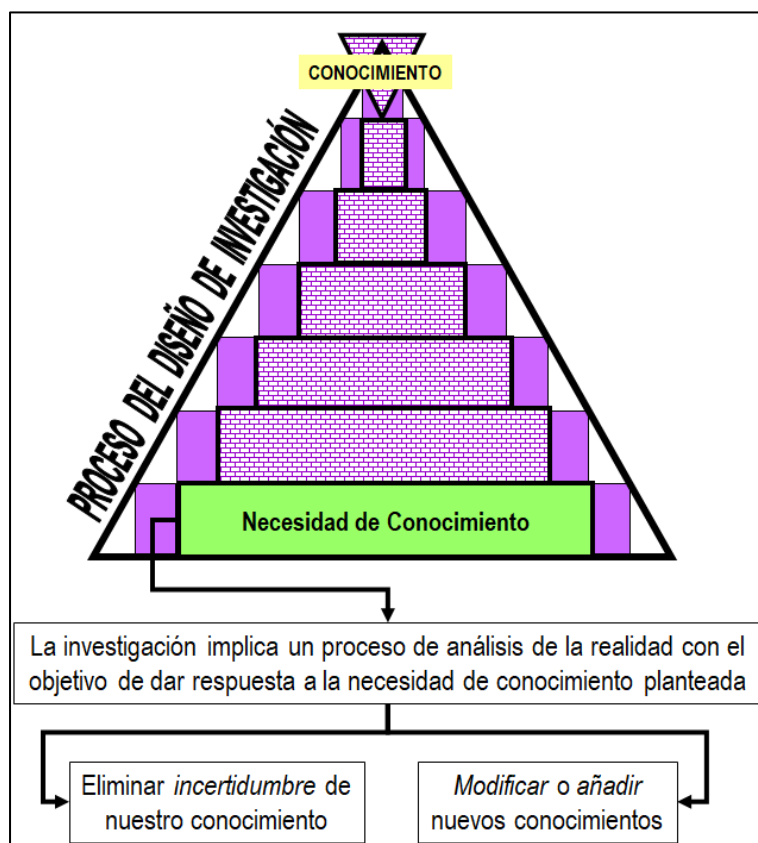


Figura 5. Necesidad de conocimiento

El problema de investigación o necesidad de conocimiento se debe definir en los siguientes términos:

- Debe expresar una relación entre dos o más variables.
- El planteamiento debe ser claro, sin ambigüedades y, en lo posible, en forma de pregunta.
- Debe permitir su verificación empírica.
- El problema debe ser relevante, es decir que justifique el esfuerzo y la inversión que se gasta en su resolución.

- Debe enmarcarse dentro de una teoría, considerando que la Ciencia busca generalizaciones y no se construye con hechos aislados.

Pregunta PICO

En los estudios de intervención, la estructura de una “pregunta” clínica de investigación o planteamiento de la necesidad de conocimiento suele adoptar la forma conocida como PICO, definiendo de forma precisa (ver Figura 6):

- El tipo de *Pacientes* o problema que se valorará en el estudio (P).
- El tipo de *Intervención* cuyo efecto se quiere estudiar (I).
- El tipo de grupo de *Comparación* que se utilizará (C).
- Y el resultado (‘*Outcome*’) que se valorará (O).

La evaluación de la calidad de la pregunta de investigación se realiza reflexionando sobre sus características y valorando que sea una pregunta factible, interesante, novedosa, ética y relevante (FINER, *Feasible, Interesting, Novel, Ethical* y *Relevant*).

| PREGUNTA CLÍNICA: | |
|--------------------------|---|
| P – | Qué tipo de P acientes o problema se valora |
| I – | Qué tipo de I ntervención se evalúa (preventiva, diagnóstica...) |
| C – | Cuál es el grupo de C omparación |
| O – | Qué resultado se evalúa (O utcome) |

Figura 6. Estructura de la pregunta clínica de investigación

La evaluación de la pregunta clínica permite redefinirla si fuese necesario, ya que no se habrían representado adecuadamente las características anteriores de FINER.

Es recomendable tener en cuenta los elementos de la pregunta PICO para elaborar un buen “título” del estudio realizado, aportando en dicho título todos los detalles que son interesantes para que el lector o lectora tenga con una información muy breve expresada en dicho título una idea lo más completa posible del

contenido que se puede encontrar si decide leer el artículo o informe de investigación.

Conocimiento previo

Una vez ya se ha planteado la necesidad de conocimiento (pregunta de investigación), para construir la investigación es necesario que el investigador o investigadora desarrolle en primer lugar, tareas de búsqueda y revisión de la información previa (*'conocimiento previo'*) para posteriormente valorar críticamente la calidad de las evidencias o pruebas aportadas en la literatura y utilizar en su estudio aquella información que ha pasado el filtro de la calidad de su evidencia. En este punto es muy importante que el investigador o investigadora y el lector o lectora dispongan de competencias de lectura crítica o activa fundamentadas en el modelo de la Práctica Basada en la Evidencia.

La literatura es el repositorio escrito del conocimiento. Su revisión ayuda a perfilar de manera más elaborada la hipótesis de investigación. Como ya se ha comentado, una buena hipótesis debe ser factible o abordable empíricamente, interesante, novedosa, ética y relevante para generar conocimiento científico.

Delimitar de forma clara, precisa y concreta el problema que se debe resolver (“necesidad de conocimiento”) permite iniciar el proceso de búsqueda y localización de la información (“conocimiento previo”) que facilitará la respuesta más adecuada, precisa y actual al problema planteado.

La revisión del conocimiento previo (revisión de la literatura) puede proceder del análisis de experiencias anteriores de investigación o estudios piloto y/o de la revisión bibliográfica de información (estudios primarios y estudios secundarios como las revisiones sistemáticas y los trabajos de meta-análisis o los estudios de meta-investigación) (ver Figura 7).

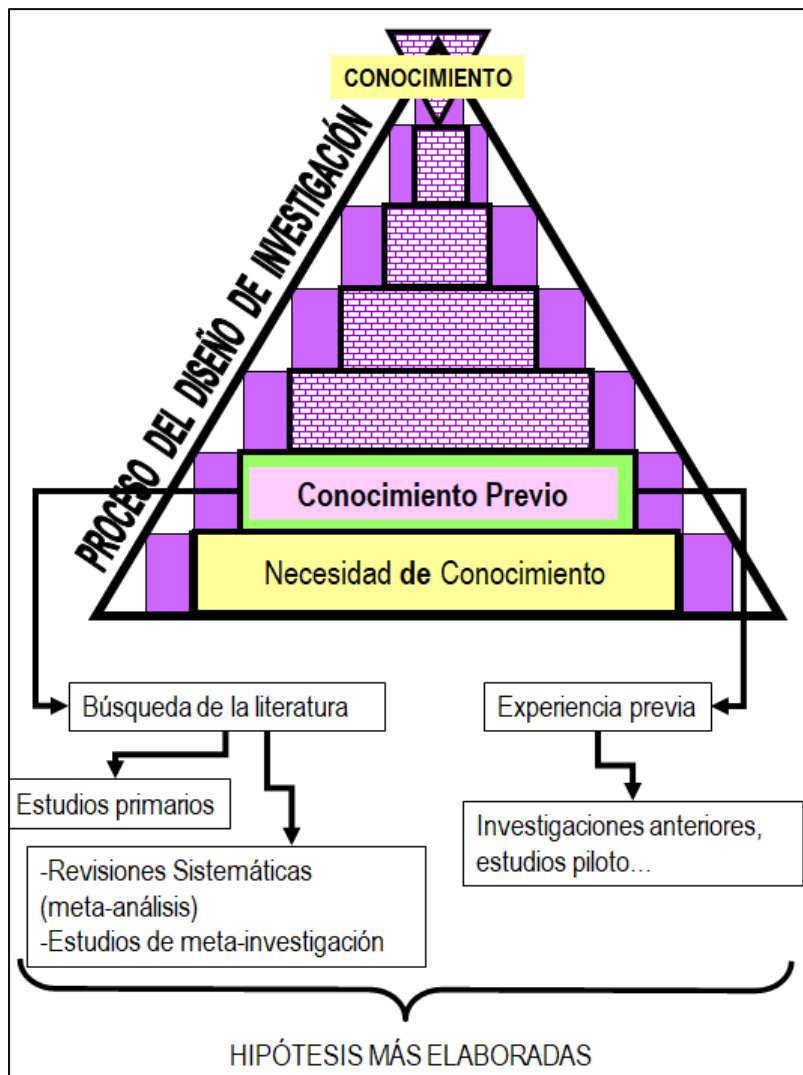


Figura 7. Conocimiento previo

Por lo tanto, una vez que se ha delimitado el problema de investigación es necesario iniciar un proceso de búsqueda de información y de valoración de dicho conocimiento previo que existe sobre el fenómeno objeto de estudio.

Actualmente, y dada la cantidad de información a la que se puede recurrir, exige utilizar una técnica de búsqueda bibliográfica que sea sobre todo metódica: definir la cuestión, seleccionar la fuente o las fuentes de información, formular el perfil de búsqueda con las palabras claves y ejecutar la búsqueda. Si los resultados no son satisfactorios entonces será necesario modificar el perfil de búsqueda o las palabras claves y/o las bases de datos seleccionadas. Las bases de datos (por ejemplo, WOS, PubMed, PsycInfo...) son una herramienta muy valiosa para obtener el conocimiento previo.

Después de la evaluación de la información obtenida con la revisión de la literatura se continúa el proceso de diseño con otras cuestiones metodológicas como delimitar las variables que van a ser investigadas y que formarán parte de las hipótesis de investigación, plantear las posibles amenazas a la validez de los hallazgos y planificar su control ya sea con el propio diseño de investigación (por ejemplo con la técnica de control de la aleatorización, con la técnica de la constancia, con el diseño de bloques ...) o con la herramienta estadística de diseño más apropiada (por ejemplo con diseños con variables covariadas). Además, también se valorará el tipo de metodología de investigación que puede ofrecer la mejor respuesta a los objetivos de investigación. La respuesta podría ser aplicar un diseño cuantitativo donde se aplique la metodología experimental, la cuasi-experimental o la metodología no experimental, un estudio cualitativo, o, quizás, un estudio con un método mixto donde se combine una parte cuantitativa con una cualitativa. Se trata de cubrir las fases de técnica y análisis metodológico del proceso de diseño de investigación descritos anteriormente en la pirámide del proceso del diseño de investigación.

Las fuentes de información que contienen el conocimiento previo pueden ser fuentes *primarias* (por ejemplo un artículo de revista, una conferencia, una tesis doctoral...) y fuentes *secundarias* donde se recogen y organizan los datos de la literatura primaria (bases de datos bibliográficas, base de datos de libros con ISBN, base de datos TESEO de tesis doctorales españolas, estudios de meta-análisis, estudios de meta-investigación...). Los artículos, libros, tesis, conferencias... son ya tan numerosos y su producción es tan rápida que se convierte en una ardua labor revisar y evaluar la información.

Las bases de datos y el acceso a Internet son una gran ayuda, pero al mismo tiempo favorecen la acumulación de información que muchas veces se queda en el cajón (en el disco duro del ordenador) sin llegar a ser leída nunca. Además, no toda la información tiene el mismo valor científico, pues tal y como ya se ha comentado la calidad de la evidencia aportada por un estudio se puede jerarquizar. Por lo tanto, existe una jerarquía en la calidad de las pruebas aportadas por las investigaciones. Por ello, es necesario evaluar de forma crítica o activa la calidad de los hallazgos de las publicaciones e informes de investigación dado que la

evidencia o las pruebas aportadas por el estudio tienen diferentes grados de validez.

Es muy importante que los programas de formación universitaria desarrollen este tipo de competencias entre los alumnos y alumnas, formando profesionales que sepan acceder a la información con las nuevas tecnologías, pero que también conozcan cómo valorar la información y jerarquizarla en función de su calidad y rigurosidad y cómo integrar la información obtenida en su tarea profesional. Competencias que el movimiento de la Práctica Basada en la Evidencia destaca como esenciales para el profesional actual.

El término de Práctica Basada en la Evidencia procede del modelo de Medicina Basada en la Evidencia cuya definición clásica señala que se trata de la utilización consciente, explícita y juiciosa por parte del profesional de la mejor evidencia clínica disponible hasta el momento, para tomar decisiones sobre el cuidado de los pacientes. Por lo tanto, integra la maestría clínica individual de los profesionales (juicio clínico) con las mejores evidencias científicas externas disponibles a partir de una investigación. Este último componente relacionado con la identificación de la evidencia o las pruebas parece, en teoría, relativamente fácil de adquirir, pero en la práctica los profesionales se ven desbordados por una cantidad de información imposible de manejar en muchas ocasiones. Hay muchas fuentes de publicación o revistas, además su calidad es muy desigual, sus contenidos se pueden quedar obsoletos en un lapsus de tiempo breve y se publica muchísima información cada día, agravándose la situación con el hecho de que el tiempo disponible para la búsqueda de información y el estudio es cada vez más limitado (Shaughnessy, 2009). Como señala Shaughnessy (2009), *“cada año los médicos deben decidir cuál de los miles de artículos recién publicados podrá tener tiempo para leer. Para determinar qué artículos son los más útiles clínicamente, los médicos deben evaluar su relevancia, validez e importancia clínica. El uso de estos criterios puede reducir drásticamente la cantidad de artículos que los médicos necesitan leer”* (p. 668). Como consecuencia de lo anterior, los conocimientos del profesional se deterioran rápidamente tras finalizar su graduación universitaria. Como señalan Gisberta y Bonfill (2004), este hecho está magníficamente representado en la gráfica clásica donde el eje de ordenadas representa los conocimientos actualizados sobre el mejor tratamiento y el eje de abscisas los años desde la

graduación (ver Figura 8). La representación gráfica señala lo que se ha denominado una peligrosa “pendiente resbaladiza” dada la relación negativa entre las variables de conocimientos actualizados y años que han pasado desde la graduación universitaria.

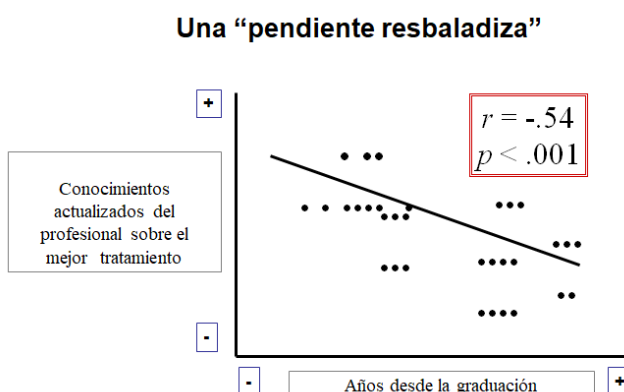


Figura 8. Relación entre conocimientos actualizados y años que han transcurrido desde la graduación

Cuando se planifica una investigación, una adecuada revisión y evaluación de la información disponible en los informes y trabajos de investigación permitirá plantear hipótesis más elaboradas, mejorando el proceso de diseño de investigación y la acumulación de conocimiento válido. En este punto, las bases de datos y la consulta vía Internet han ayudado de forma considerable a la tarea de revisión de la literatura y actualmente la dificultad no estriba tanto en obtener información sobre una temática sino más bien en cómo valorar de forma crítica la calidad de las evidencias que aportan los estudios. Y esa es una competencia indispensable del investigador o investigadora moderno y del profesional como consumidor de literatura científica.

Las denominadas “listas guías de comprobación” o “guías de publicación” son una gran ayuda, tanto para que el investigador o investigadora chequee o compruebe la calidad de los resultados de su estudio antes de enviar el manuscrito a la revista para su futura publicación como para que los revisores y las revisoras de las revistas tengan un protocolo de comprobación del proceso de diseño de investigación llevado a cabo por el autor del artículo. Por ejemplo, destacan las guías denominadas (González de Dios, González-Muñoz, Alonso-Arroyo, y Benavent, 2014):

- JARS (Journal Article Reporting Standards), elaborada desde el grupo de trabajo de la American Psychological Association (JARS Group) (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; Appelbaum, Cooper, Kline, Mayo-Wilson, Nezu, & Rao, 2018; Levitt, Bamberg, Creswell, Frost, Josselson, & Suárez-Orozco, 2018).
- CONSORT (CONsolidated Standards fOr Reporting of Trials) para ensayos clínicos aleatorios o estudios realizados con una metodología experimental.
- TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) para estudios con una metodología cuasi-experimental sin asignación aleatoria del tratamiento.
- STROBE (STrengthening the Reporting of Observational studies in Epidemiology) para estudios no experimentales u observacionales.
- STRAND (STandards for Reporting of Diagnostic Accuracy) para estudios de precisión diagnóstica.
- PRISMA para valorar la calidad de las revisiones sistemáticas y los estudios de meta-análisis.
- Los listados CASPE (Critical Appraisal Skills Programme Español) para realizar lecturas críticas de estudios con diferentes metodologías de investigación (<http://www.redcaspe.org>). CASP International es una organización que proporciona materiales para la lectura crítica de la evidencia clínica como plantillas o listados de comprobación (listas de verificación) para evaluar diferentes tipos de evidencia (ensayos clínicos, revisiones sistemáticas, estudios de diagnóstico, estudios cualitativos, estudios de predicción clínica, estudios de casos y controles, estudios de cohorte y análisis de evaluaciones económicas) y calculadoras de determinados estadísticos.

También los consumidores y las consumidoras de literatura o lectores y lectoras pueden hacer uso de dichas guías para realizar una lectura crítica de los trabajos de investigación, chequeando y valorando cada uno de sus puntos cuando realiza

la lectura de un artículo o informe. En todos los casos, las listas de comprobación facilitan una valoración crítica de los hallazgos de una investigación, enmarcando la conducta del profesional y del investigador o investigadora en el modelo de 'Práctica Basada en la Evidencia'.

En resumen, el objetivo de la fase de elaboración del conocimiento previo es revisar de forma sistemática la documentación que existe sobre el tema y esto exige definir un protocolo de consulta que incluya los repertorios bibliográficos o las bases de datos más relevantes dentro de la temática de investigación (fuentes de información secundaria). Una de las bases de datos más utilizada es MEDLINE de la *National Library of Medicine* junto con su versión PubMed en Internet. Dentro del área médica destaca también la base de datos Excerpta Medica (denominada EMBASE). En Psicología sobresale la base de datos PsycInfo, producida por la *American Psychological Association* (APA) y en Educación se encuentra ERIC (*Education Resources Information Center*). Una plataforma de reconocido prestigio de búsqueda multidisciplinaria que permite hacer uso de varias bases de datos de forma simultánea en una misma interface es Web of Science (WoS). A partir de ella se elaboran los conocidos factores de impacto de las revistas (JCR, Journal Citation Report) y los índices de citas de los artículos. Por otra parte, una fuente de información bibliográfica primaria de gran importancia es la revista de *Annual Reviews* (<http://www.annualreviews.org/>). El *Annual Reviews* es un conjunto de revistas especializadas en una temática como la Psicología (*Annual Reviews of Psychology*), la Psicología Clínica (*Clinical Psychology*), la Estadística y sus Aplicaciones (*Annual Reviews of Statistics and its Application*), la Economía (*Annual Reviews of Economics*), la Sociología (*Annual Reviews of Sociology*), la Criminología (*Annual Reviews of Criminology*), la Salud Pública (*Annual Reviews of Public Health*) o la Medicina (*Annual Reviews of Medicine*), por ejemplo.

Conviene tener en cuenta que las bases de datos informatizadas no consiguen localizar todos los trabajos publicados ni están completamente actualizadas, ya que se requiere un tiempo desde la publicación del artículo hasta la introducción de la información en la base. Las bases de datos tampoco son exhaustivas. La denominada 'literatura gris', es decir, aquella que se publica en resúmenes de congresos, informes de organismos o instituciones... no aparecen en dichos repertorios. Por lo tanto, es conveniente completar la búsqueda por otras vías como

el contacto directo con el investigador experto en el tema o la consulta directa de la documentación que suelen ofrecer los centros de investigación en sus páginas Web.

Actualmente el servicio que ofrecen las revistas de alertar al usuario vía correo electrónico cuando se produce la aparición de un nuevo número de la revista es una de las mejores opciones para maximizar el acceso rápido a los últimos avances que se producen en el campo científico de interés. También, el servicio de alertas de citación de un artículo que proporcionan las bases de datos es realmente muy útil.

En definitiva, actualmente la revisión del conocimiento es una ardua tarea que el investigador debe realizar, sobre todo, de forma sistemática con el objetivo de representar las diferentes perspectivas de los autores y el desarrollo de la temática que se desea estudiar. Hay que evitar una búsqueda de la información sesgada ya sea por los intereses del investigador (solamente le interesan las publicaciones que apoyan una teoría por ejemplo, o las que siguen un determinado modelo), o por acceder solamente a la literatura que está disponible de forma sencilla para el investigador como la información depositada en una biblioteca o en su centro de trabajo. Ciertamente, las revisiones sistemáticas y los trabajos de meta-análisis representaron un hito importante en la formación teórica sobre un fenómeno, pero de nuevo la lectura crítica y activa de este tipo de trabajos, como el de los estudios primarios, es necesario para seguir filtrando la calidad de los hallazgos publicados. El mismo control de calidad y jerarquía de la evidencia que se aplica cuando se lee un estudio primario debe ser aplicado cuando se lee un estudio de revisión sistemática o de meta-análisis.

A las bases de datos, los estudios primarios encontrados en las revistas, los informes de las asociaciones e instituciones y la literatura gris, en general, hay que añadir ahora los denominados pre-prints o documentos preimpresos que día a día inundan de nuevo la información que hay sobre un tema y los pre-registros de los estudios. Conviene estar muy alerta cuando se lee un trabajo de pre-print depositado o almacenado en un repositorio, ya que dichos trabajos aún no han pasado por el filtro de la revisión por pares que es indispensable para publicar un artículo en una revista científica (revisión crítica por parte de otros investigadores

como fase previa para que un manuscrito sea aceptado para su publicación en una revista) y, por lo tanto, la calidad de su diseño de investigación no ha sido valorada.

Esa revisión por pares o por expertos de los manuscritos enviados para su posible publicación tiene como objetivo identificar los puntos débiles del diseño del estudio y puede conducir a la decisión de rechazar el manuscrito, proponer cambios importantes para mejorar la calidad de la presentación y redacción del estudio, proponer cambios menores o directamente su publicación (este último hecho se produce en muy pocas ocasiones). Por supuesto, este filtro científico es fundamental para obtener conocimiento válido, pero no es garantía absoluta de que los artículos carezcan de errores graves. En la última década ha aumentado la tasa de artículos retirados de las revistas, una vez publicados, ya que se ha demostrado, por el análisis de meta-investigación o por la revisión de los lectores o lectoras expertos, que habían errores que no se detectaron en el proceso de análisis por los pares o expertos antes de su publicación. Esta es una de las razones que aconseja comenzar la lectura de un artículo con una postura crítica de análisis del proceso del diseño en todas sus fases, desde la revisión del conocimiento previo (calidad y actualización de la información presentada en el apartado de introducción) hasta la interpretación de los resultados y el conocimiento teórico que se aporta en el apartado de discusión. Dicha lectura exige activar las competencias de análisis activo y los conocimientos del área de la metodología de investigación. Si hay carencias en la formación metodológica entonces no se podrá llevar a cabo una lectura crítica o activa eficaz. Y, además, esas competencias requieren una re-educación constante para estar actualizados en cómo se presenta y se redactan los resultados científicos en cada contexto de investigación. El lector o lectora es el responsable para decidir qué valor tiene el artículo o informe que lee.

Ya se ha comentado que el auge de la Ciencia en Abierto probablemente impulsará cambios destacados en la forma de hacer, leer y consultar la información. Los deseos de transparencia que demanda la Ciencia en sus informes y trabajos para facilitar la observación, el control de la calidad de sus hallazgos, evitando los falsos positivos y la mala interpretación de los resultados nulos (cuando se mantiene la hipótesis nula) y que la cualidad auto-correctiva que tiene la Ciencia realmente funcione exigen un cambio en todos los agentes implicados en el desarrollo de los hechos científicos. Los agentes incluyen a los mismos científicos y

su práctica de investigación, a las políticas de los editores de las revistas, a las instituciones, al software estadístico, a la ética del investigador y su re-educación en el proceso del diseño de investigación que afecta de forma directa dentro del mundo académico, a las enseñanzas que reciben los alumnos y alumnasy, también, entre los lectores y lectoras de material científico que a pesar de no construir investigaciones necesitan saber y comprender si aquello que leen se construyó con fiabilidad y validez. La sociedad del conocimiento avanza si el conocimiento está sometido al proceso riguroso del método científico.

Hipótesis de investigación

Después de la evaluación de la información obtenida con la revisión de la literatura se continúa el proceso de diseño con otras cuestiones metodológicas como delimitar las variables que van a ser investigadas y que formarán parte de las hipótesis teóricas o sustantivas de la investigación y su operacionalización, así como plantear el control de otras variables cuyos efectos deben mantenerse constantes en el diseño para evitar que contaminen los resultados.

La hipótesis de investigación es la denominada hipótesis científica o hipótesis sustantiva del estudio.

Para poder completar el proceso del método científico es necesario operacionalizar la hipótesis científica en un enunciado contrastable empíricamente (hipótesis científica operacionalizada) y así pasar a la fase de 'contrastación estadística' con datos empíricos del estudio. En esta fase de contrastación estadística mediante los procedimientos clásicos de contraste de hipótesis se formulan las dos hipótesis estadísticas: la *hipótesis nula* (hipótesis de nulidad del efecto, que generalmente plantea un efecto de cero del tratamiento) y la *hipótesis alternativa* (que plantea, generalmente, un efecto diferente a cero).

Planificación de la investigación

Llegados a este punto del proceso de diseño de investigación es necesario detenerse y planificar de forma cuidadosa el diseño del estudio donde se reflexionará sobre las posibles amenazas a la validez interna de los resultados (así como prestar atención al resto de tipos de validez: de conclusión estadística, de

constructo y externa) y, si es el caso, se planificará su control ya sea con la propia metodología del estudio como por ejemplo la asignación aleatoria del tratamiento en la metodología experimental (técnica de la aleatorización), con técnicas concretas de control como la constancia o la eliminación de la variable contaminadora, el apareamiento y/o con la factorización de la variable contaminadora en la propia ecuación estructural del diseño (por ejemplo, los diseños de bloques o los diseños de cuadrado latino) o con la herramienta estadística de diseño más apropiada (por ejemplo los diseños con variables covariadas).

En la fase de planificación del estudio también es necesario reflexionar sobre la calidad de la medida de los constructos (fiabilidad y validez) que van a ser estudiados (validez de constructo), sobre el grado de generalización de los hallazgos a otras situaciones y momentos temporales (validez externa) y sobre la calidad del proceso de inferencia estadística que se realizará (validez de conclusión estadística), siendo especialmente relevante en este último punto planificar el tamaño de la muestra en función de los criterios de alfa, potencia estadística deseada y tamaño del efecto esperado así como la comprobación de los supuestos estadísticos que las pruebas estadísticas requieren para que funcionen de forma adecuada y no estén sesgadas por artefactos metodológicos que podrían dañar la calidad de sus resultados (validez).

Por lo tanto, una fase de especial relevancia dentro del proceso del diseño de la investigación es la etapa de la planificación de los elementos que forman parte del diseño del estudio. Durante la planificación de la investigación, el investigador toma una serie de decisiones sobre:

- 1- La elección de las variables que forman el experimento y que operacionalizan a las variables constructo objeto de estudio.
- 2- La estrategia de recogida de datos.
- 3- El tipo de diseño más adecuado para contrastar las hipótesis estadísticas planteadas, controlando las posibles variables que podrían contaminar los resultados.

Se trata de los tres elementos básicos que darán forma al diseño de la investigación que se aplicará en un determinado estudio.

En opinión de Kerlinger (1986), el principal propósito de la fase de planificación del estudio es asegurar el objetivo metodológico conocido como *MAX-MIN-CON* que está relacionado con la variabilidad que se produce en la variable dependiente:

- *MAXimizar la varianza sistemática primaria* de la variable dependiente: vinculada al efecto de la variable independiente de tratamiento.
- *MINimizar la varianza del error* de la variable dependiente: variabilidad debida al azar, errores de medidas o provocados por diferencias aleatorias relacionadas con los propios sujetos, la situación experimental o el medio ambiente.
- *CONtrolar la varianza sistemática secundaria* de la variable dependiente: se genera por el influjo de alguna variable extraña relevante que actúa de forma sistemática por sí sola o junto con la variable de tratamiento.

La “varianza sistemática” es aquella desviación o variabilidad que presentan los datos (puntuaciones en la variable dependiente) hacia una dirección más que otra. Es decir, los datos varían de forma sistemática en un determinado sentido. Y la varianza sistemática que se produce en los datos (variable dependiente) puede ser primaria (provocada por la variable independiente cuyo efecto se desea estudiar en la investigación) o secundaria (provocada por las terceras variables o variables extrañas cuyo efecto provoca variabilidad que contamina a los resultados).

La varianza de error es la variabilidad que presentan las puntuaciones en la variable dependiente debido a factores aleatorios. No es una variabilidad sistemática sino que se trata de una varianza aleatoria. Este tipo de variabilidad suele deberse a errores de medida debido a la utilización de instrumentos que no tienen unas buenas propiedades psicométricas o están mal calibrados, a factores relacionados con los propios individuos (diferencias individuales), a la propia situación experimental o a las condiciones ambientales. La varianza del error se conoce como variabilidad intra-grupo o variabilidad intra-tratamiento, ya que diferencia a los sujetos o a las unidades experimentales a pesar de que puedan recibir el mismo tratamiento o pertenezcan al mismo grupo. Por lo tanto, dado que

se trata de una varianza aleatoria no se puede controlar, pero el investigador o investigadora sí puede minimizarla o reducirla a través de un diseño de investigación que vigile posibles fuentes aleatorias de error.

El diseño más eficaz es aquel que controla la varianza sistemática secundaria, extrayéndola de la varianza total de la variable dependiente de modo que se pueda comparar la varianza sistemática primaria y la varianza del error (aleatoria) cuando se ejecuta una prueba estadística sin la presencia de variables extrañas que contaminen la verdadera variación de las observaciones, maximizando de este modo el efecto de la varianza sistemática primaria. La aplicación de los diseños de investigación que posibiliten el principio *MAX-MIN-CON* permitirá obtener un conocimiento válido del fenómeno planteado en las hipótesis de investigación.

Método

En la fase de método se desarrolla el diseño de investigación planificado previamente (participantes, tamaño de la muestra, instrumentos de medida, procedimiento llevado a cabo para recoger los datos y para poder ejecutar el experimento) y se valora qué tipo de metodología de investigación ofrecerá la mejor respuesta a los objetivos e hipótesis de investigación y a la estrategia de recogida de los datos: metodología experimental, cuasi-experimental, no experimental o $N = 1$. O quizás se llegue a la decisión de optar por una metodología cualitativa o, quizás, lo más adecuado sea aplicar una metodología mixta donde se combiene una parte de análisis cuantitativo y otra de análisis cualitativo. Estas cuestiones serán tratadas en profundidad cuando se analicen las diferentes metodologías de investigación.

Análisis de los datos

Siguiendo con los pasos ejemplificados en la pirámide del proceso de investigación, el análisis de los datos se ajustará a los objetivos de las hipótesis de investigación y a su planificación gracias al contraste de las hipótesis estadísticas mediante la prueba estadística que se considere más adecuada, acompañada de información sobre la magnitud del efecto detectado (estadístico del tamaño del efecto y su intervalo de confianza) así como de otros análisis que podrían ser

relevantes (por ejemplo, el valor del Factor Bayes (FB) en los contrastes de hipótesis) y la interpretación clínica o sustantiva de los hallazgos.

Conviene tener muy presente que en los procedimientos habituales de análisis de datos basado en el modelo de comprobación de la significación de la hipótesis nula (Null Significance Statistical Testing, NHST), el contraste de hipótesis se inicia siempre asumiendo que la hipótesis nula es cierta (y se conoce su distribución) y sobre dicho supuesto se obtiene la probabilidad del resultado obtenido en la distribución de la hipótesis nula. Posteriormente, el tema del contraste estadístico será tratado con detalle, ya que su comprensión y uso adecuado son competencias básicas del investigador o investigadora y también del lector o lectora consumidores de literatura científica.

Conocimiento adquirido

Finalmente, los resultados de los análisis estadísticos aportarán unos hallazgos que serán debatidos o discutidos teóricamente dentro del campo científico donde se desarrolla la investigación, dando lugar a un nuevo conocimiento científico que pasará a formar parte del conocimiento previo de un futuro estudio. Así, una vez obtenidos los resultados, la teoría propuesta en el estudio es aceptada o modificada, incorporando las nuevas evidencias detectadas por deducción y se iniciará de nuevo el proceso de investigación científica, formulando vía inducción suposiciones teóricas aceptables.

En este punto es muy importante destacar otras competencias básicas de los investigadores y las investigadoras relacionadas con la redacción cuidadosa de los resultados y con la valoración correcta del alcance de los resultados, así como de sus limitaciones o problemas que podrían hacer cambiar los hallazgos o moderar el alcance de su generalización. También los lectores y las lectoras deberán valorar de forma activa (lectura crítica) la calidad de la redacción de los resultados y la calidad de la discusión y conclusiones que se detallan en el artículo o informe.

En resumen, la Psicología es una ciencia que construye enunciados contrastables empíricamente acerca del comportamiento humano y el método tiene la finalidad general de contrastar dichos enunciados o construcciones teóricas con los resultados de sus comprobaciones como un proceso sistemático de validación.

La Psicología es una ciencia más cuyo método de conocimiento es el propio de la Ciencia en general: el método hipotético-deductivo que muy esquemáticamente supone formular un enunciado general (es decir, una hipótesis) y contrastarla con la realidad. Así, el método de la investigación científica supone un proceso iterativo de razonamiento inductivo y razonamiento deductivo (Nesselroade y Cattell, 1988), donde a partir de la observación de ciertos hechos se infiere por inducción suposiciones teóricas aceptables que expliquen cierta regularidad de los mismos que a su vez permiten derivar consecuencias por deducción que podrán ser contrastadas con los datos. Si el procedimiento es satisfactorio entonces el enunciado se considera válido, pero si no se considera satisfactorio entonces el enunciado debe reformularse. Se trata de la conocida espiral del método hipotético-deductivo donde la inducción representa el comienzo del proceso científico con una teoría inicial acerca de la naturaleza de los datos.

El desarrollo de la Ciencia es siempre progresivo a través de un proceso cíclico de aprendizaje guiado (Box, Hunter, y Hunter, 1978, 2005). La unión entre el diseño de la investigación y la estadística es necesaria; la estadística constituye la “*tecnología del método científico*” (Mood y Graybill, 1972) y el procedimiento de comprobación de hipótesis estadísticas requiere trabajar con un diseño de la investigación que garantice que las conclusiones que se obtengan no estarán invalidadas por factores no controlados en el diseño o por sesgos que amenazan la validez de los resultados del estudio.

Por lo tanto, un diseño de investigación es una estrategia de estudio que supone manipular o seleccionar variables, medir variables y también controlar variables de sesgo para poder encontrar una respuesta lo más cercana posible a la realidad de la cuestión de investigación. Así, es necesario que el investigador tenga en cuenta en la fase de planificación de su estudio las posibles variables extrañas que podrían afectar a la calidad de los hallazgos ya sea para controlar su efecto, para reducir la varianza del error de las puntuaciones (en el diseño debe ser varianza de error aleatorio) o para provocar ambas cosas a la vez.

La validez interna del estudio (la calidad de los resultados) sólo podrá estar garantizada si las variables extrañas son variables controladas con el diseño de investigación, manteniendo constante su efecto sobre la relación entre las variables

que forman la hipótesis de investigación y que se conocen como las 'variables explicativas' del modelo (variable independiente y variable dependiente).

Realizar una adecuada revisión del conocimiento previo permite descubrir posibles variables extrañas cuyo efecto sistemático distorsionaría los resultados y dicha revisión posibilita planificar su control en el diseño. Por ello, la reflexión crítica sobre la documentación que aporta el conocimiento previo es una competencia del investigador o investigadora que favorece la planificación de la investigación y evita sesgos sistemáticos que amenazan la validez de los hallazgos.

Capítulo 4. Metodologías de investigación

Dolores Frías-Navarro
Universidad de Valencia

Citar el capítulo como:

Frías-Navarro, D. (2020). Metodologías de investigación. En D. Frías-Navarro y M. Pascual-Soler (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP>
Disponible en: osf.io/kngtp

Toda investigación debe estar adecuadamente diseñada, eficientemente ejecutada, correctamente analizada, bien interpretada y claramente presentada y redactada. Todas estas competencias requieren un gran esfuerzo por parte de los investigadores y las investigadoras y la calidad de los hallazgos o resultados (su grado de validez) está en gran medida relacionada con la metodología de investigación que se ha podido aplicar en el estudio y con el diseño que se ha planificado junto a su correcto desarrollo, análisis e interpretación.

Como ya se ha comentado, el proceso de la investigación científica (método científico) comienza porque el investigador o investigadora se plantea un problema concreto ('necesidad de conocimiento') que debe resolver ofreciendo una explicación del fenómeno lo más válida posible o, quizás, porque plantea una nueva explicación de los hechos que requiere recoger datos para avanzar en esa línea de investigación o, tal vez, podría tratarse de la necesidad de replicar un estudio o de reproducir los resultados de una investigación. Este problema le conduce a elaborar un plan de investigación cuyo objetivo principal es obtener observaciones y datos

relevantes y válidos para la hipótesis empírica planteada que operacionaliza a la hipótesis científica o sustantiva del estudio.

El plan de investigación (la planificación del diseño del estudio) requiere que el investigador tome decisiones respecto a la estrategia de recogida de datos que depende, fundamentalmente, de aspectos metodológicos y técnicos como “*la forma de operativizar las variables de la hipótesis, la posibilidad o no de manipular la variable independiente, la capacidad de controlar las variables extrañas y de confundido, el grado de selección aleatoria de las unidades de observación (por lo general, sujetos), su asignación a las diferentes condiciones o niveles de actuación de la variable independiente, la estructuración interna del procedimiento, etc.*” (Arnau, 1989, p. 585).

Una vez resueltos estos aspectos, y muy especialmente la posibilidad o no de asignar aleatoriamente los niveles o condiciones de la variable independiente a los grupos de estudio y la de su manipulación, el diseñador o diseñadora de la investigación debe entonces “seleccionar la estrategia más adecuada para obtener los datos”. Es decir, se trata de determinar la metodología del estudio así como los sistemas de registro para la medida de la variable dependiente. Son los dos aspectos clave del proceso del diseño de investigación. Por lo tanto, como consecuencia del planteamiento de la solución del problema, el investigador o investigadora tendrá que utilizar la modalidad de investigación (metodología) que sea más adecuada o posible para solucionar el problema planteado (necesidad de conocimiento).

Por lo tanto, el método científico de investigación como procedimiento general de obtener información es único y el cumplimiento de sus principios permite alcanzar el conocimiento científico, pero cada problema de investigación requiere una determinada actuación científica que hace que exista una pluralidad de modalidades o métodos de investigaciones vinculados directamente con la elección de la estrategia de recogida de datos. Metodologías que pueden ser cuantitativas, cualitativas o mixtas dado que se produce una combinación de análisis cuantitativo y cualitativo.

Para poder contrastar adecuadamente la validez de los enunciados propuestos en un estudio, los métodos (metodologías) de investigación tienen que garantizar

que los procedimientos que se aplican aseguran dicho fin, no siendo indispensable que utilicen procedimientos matemáticos, aunque su uso puede reducir las ambigüedades a la hora de la contrastación con los datos así como clarificar la relación entre las variables.

Diferencias entre las metodologías de investigación

Pero, ¿qué diferencia a las distintas metodologías de investigación cuantitativa? Las diferencias entre las metodologías de investigación no es ni el contenido de estudio, ni la población de referencia que se estudia, ni el contexto en el que se estudia; las diferencias radican en la respuesta dada a dos interrogantes:

1) *Manipulación* (sí / no) de las condiciones de la variable independiente o factor de tratamiento. Es decir, manipulación (variación controlada por el investigador o investigadora, se trata de una variable activa) de las condiciones de la variable independiente o, en cambio, se trata de una variable asignada no manipulada (variable seleccionada por el investigador).

2) *Asignación aleatoria* (sí / no) de las condiciones de la variable independiente. Es decir, aleatorización en la asignación de los tratamientos a las unidades experimentales (sujetos) en los diseños de medidas independientes o entre-sujetos (entre-grupos) o, quizás, aleatorización en el orden de presentación de las condiciones de la variable objeto de estudio (medida) en los diseños de medidas repetidas o intra-sujetos.

Hay que destacar que la denominada metodología *experimental* es la única que incluye ambos requisitos. La investigación *cuasi-experimental* únicamente cumple el requisito de manipulación de la variable independiente y la metodología *no experimental* (diseños de encuesta y los diseños observacionales) no cumple ninguno de los dos.

Los *verdaderos diseños experimentales* (en términos de Campbell y Stanley, 1966) son los experimentos aleatorizados, es decir aquellos en los que además de la manipulación de la variable independiente existe una asignación aleatoria de los sujetos a las condiciones de tratamiento.

Por lo tanto, las características de 1) manipulación o no manipulación de las condiciones de la variable independiente (tratamiento) y 2) la posibilidad de la asignación aleatoria del tratamiento a las unidades experimentales (generalmente son sujetos en los diseños entre grupos o la asignación aleatoria en el orden de administración del tratamiento en los diseños de medidas repetidas) son los dos elementos clave que identifican la naturaleza de la metodología empleada en el diseño de una investigación y la naturaleza causal o no de las relaciones encontradas entre las variables (ver Figura 9).

Nos encontramos en el apartado de Método de investigación en la pirámide del proceso del diseño de investigación donde se describe la metodología del estudio: *experimental, cuasi-experimental o no experimental*.

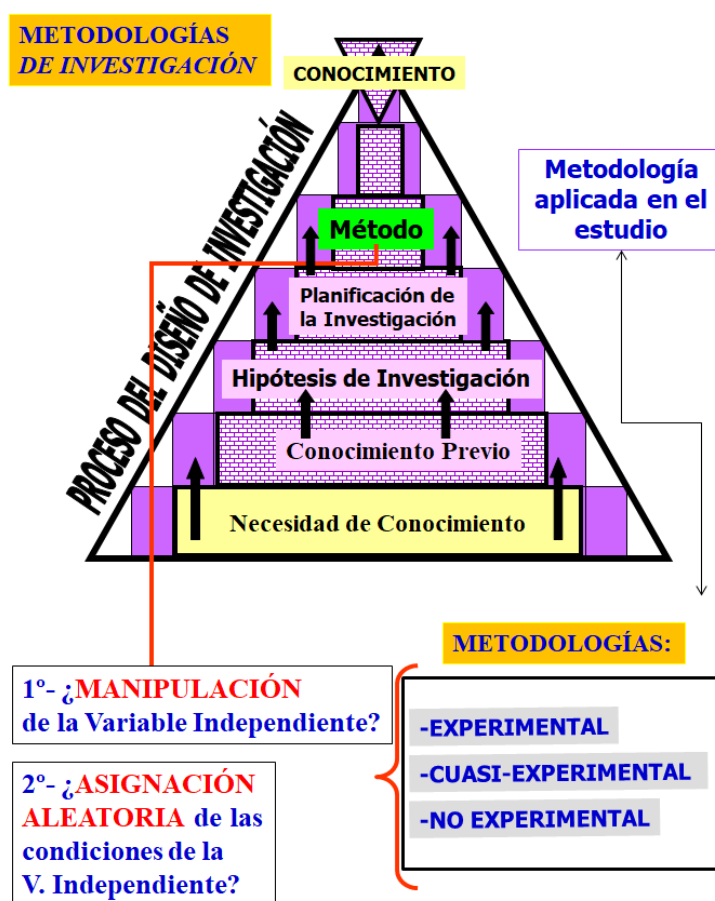


Figura 9. Metodologías de investigación

La conocida subdivisión de Campbell y Stanley (1966) clasifica las metodologías de investigación en experimentales, cuasi-experimentales y correlacionales. Correlacional no en el sentido analítico como técnica estadística

sino como búsqueda de relaciones de covariación o asociación entre las variables, donde la variable independiente es únicamente observada y no es manipulada. Cook y Campbell (1979) introducen el término cuasi-experimental y con el trabajo de Pedhazur y Schmelkin (1991) se presenta una acertada clasificación tripartita que obvia el término correlacional, eliminando así la ambigüedad que el término conlleva al vincularlo a la técnica estadística, distinguiendo entre metodología experimental, cuasi-experimental y no experimental (incluyendo en esta última categoría a los estudios observacionales y los de encuesta). Los diseños de $N = 1$ son tratados como un método de investigación con características particulares de tamaño muestral, recogida y análisis de los datos y control de la calidad de los hallazgos.

En resumen, el criterio de clasificación de las metodologías de investigación en Psicología de Pedhazur y Schmelkin (1991) de metodología *experimental*, metodología *cuasi-experimental* y metodología *no experimental* se basa en dos cuestiones que se deben poder identificar de forma clara en el informe de investigación o artículo (Frías-Navarro, 2011):

1) Presencia o no de la manipulación de las condiciones de la variable de tratamiento o variable de intervención (variable independiente).

2) Utilización o no de la asignación aleatoria de los grupos de tratamiento o grupos de comparación que forman las condiciones del diseño de investigación a las unidades experimentales (generalmente sujetos). Es decir, asignación aleatoria del tratamiento a los sujetos en los diseños entre-sujetos. O, quizás, asignación aleatoria en el orden de la presentación de las medidas o condiciones de la variable en los diseños intra-sujetos o de medidas repetidas.

Metodologías: experimental, cuasi-experimental y no experimental

En la metodología experimental la manipulación de la variable independiente se efectúa sobre las condiciones bajo las que se generarán los datos, es decir, implica que el investigador o investigadora cree situaciones de investigación (condiciones de la variable independiente; por ejemplo configuración de las pautas de intervención del grupo experimental que recibe la intervención y elaboración de las

características del grupo de control que podría ser, por ejemplo, permanecer en una lista de espera) para observar cómo funciona la variable medida (variable dependiente; por ejemplo sintomatología depresiva) en cada condición de investigación. Además, debe existir una asignación aleatoria del tratamiento o condiciones de la variable independiente o una asignación aleatoria en el orden de presentación de las medidas repetidas.

En la tabla 1 se resumen las principales características de la metodología experimental, cuasi-experimental y no experimental y, además se sitúa al ensayo clínico controlado aleatorio como sinónimo de un diseño con metodología experimental.

Tabla 1. Metodologías de investigación

| Características | Metodología | | |
|---------------------------|---|---------------------------|--|
| | <i>Experimental</i> <i>Ensayo clínico controlado aleatorio</i> | <i>Cuasi-experimental</i> | <i>No experimental</i> (Encuesta y Observacional) |
| Manipulación de variables | Sí | Sí | No |
| Asignación aleatoria | Sí | No | No |

Por lo tanto, en los estudios con una metodología experimental, el investigador o investigadora asigna aleatoriamente las condiciones de tratamiento manipuladas a los sujetos (se conoce también como ‘*ensayo clínico controlado aleatorio* ECCA o ‘*ensayo clínico aleatorio*’, ECA).

En cambio, en los estudios con una metodología cuasi-experimental aunque sí que existe una intervención/manipulación de la variable independiente, no existe la asignación aleatoria del tratamiento a los sujetos. Es decir, en los diseños con una metodología cuasi-experimental, los grupos o condiciones de la investigación no se forman de forma aleatoria sino que un criterio no aleatorio determina la configuración de los miembros que forman la condición o grupo de participantes.

Y, en los diseños con una metodología no experimental solamente se observan las variables tal y como ocurren en la realidad bajo las condiciones que ya tienen asignadas o son propias del grupo, pues no existe manipulación / intervención sobre la variable considerada independiente y por ello no es necesario plantearse la

posibilidad de la asignación aleatoria o no del tratamiento; por ejemplo, la edad, ser hombre o mujer, la orientación sexual, el nivel socioeconómico, la personalidad o el bienestar percibido. Se puede identificar a la metodología no experimental como la metodología “NiNi”: “*Ni manipulación Ni asignación aleatoria*”. Se trataría de ‘grupos intactos’, ya que no reciben ningún tipo de intervención que les pueda modificar su conducta. En el diseño de este tipo de estudios el objetivo es registrar las características de los grupos tal y como están formados y analizar sus posibles diferencias en las hipótesis que se plantean en la investigación.

Por lo tanto, cuando se habla del experimento como ‘experimento verdadero’, es decir, en el sentido de un plan de investigación sobre la naturaleza de un cierto fenómeno cuyo diseño incluye los elementos de la metodología experimental propiamente dicha, se está infiriendo que se dan las siguientes condiciones (Pascual-Llobell, García-Pérez y Frías-Navarro, 1996):

1. La existencia de, al menos, una variable manipulada, denominada variable independiente. La manipulación supone que el investigador puede seleccionar varios valores de dicha variable, delimitando para cada valor una condición experimental distinta.

2. La asignación al azar de las unidades experimentales a las distintas condiciones experimentales en los diseños entre-grupos. Es muy importante no confundir número de observaciones (N) con número de sujetos (S), ya que únicamente es verdad en los denominados diseños entre-sujetos. En los diseños intra-grupos la condición son cada una de las mediciones que se realizan (medición 1, medición 2...) y se produce la asignación al azar en el orden de presentación de los tratamientos. Por ejemplo, en un diseño con 5 sujetos y 2 mediciones, el diseño tiene 10 observaciones ($N = 10$), siendo $S = 5$.

3. La comprobación del efecto de la manipulación de la variable independiente sobre la variable de medida, conocida como variable dependiente.

4. Además, se controla cualquier otra fuente de variación que no habiendo sido manipulada deliberadamente por el investigador puede afectar de forma sistemática a la situación experimental. Se trata de las denominadas ‘variables extrañas’ o ‘terceras variables’ que deben ser eliminadas o mantenidas constantes cuando se conocen y es posible, o aleatorizarlas con el propósito de homogeneizar su efecto

en las diferentes condiciones o grupos, convirtiéndose así en 'variables controladas' en el diseño de la investigación.

Las hipótesis causales (causa → efecto) son una clase de hipótesis muy importantes dentro de los intereses de investigación, pero no son las únicas, por ejemplo las hipótesis de asociación (covariación entre las variables) cubren un área muy amplia de la investigación (fumar y cáncer; autoestima y éxito académico; ansiedad y rendimiento académico). Cook y Campbell (1979) señalan que la relación causal indica covariación entre la variable independiente y dependiente, precedencia temporal de la variable independiente o causa y explicaciones alternativas del cambio no plausibles. Siguiendo estos requisitos, el método experimental es el único que permite plantear y comprobar hipótesis de causalidad.

Por lo tanto, para inferir que A causa B se requieren tres condiciones (Kenny, 1979):

1) Establecer la asociación estadística entre los valores de A y B de manera que cuando se da la presunta causa aparece el presunto efecto y cuando no se da la presunta causa no aparece el presunto efecto.

2) Establecer la dirección de la causalidad por ejemplo de A a B, basado por ejemplo en un criterio temporal: precedencia temporal de la causa (manipulación previa de la variable independiente de tratamiento).

3) Eliminar los efectos de todas las posibles causas comunes de A y de B (ausencia de espuriedad), lo que implica el control previo de las variables extrañas, evitando explicaciones alternativas de los hallazgos.

Eliminar esas causas comunes (variables extrañas perturbadoras o contaminadoras) es un elemento clave en la definición del tipo de metodología. Generalmente los diseños emplean el control y la aleatorización para eliminar esas causas comunes de A y B. El control experimental supone mantener los valores de las posibles causas comunes constantes en todas las condiciones de tratamiento para que afecten por igual a todos los datos del estudio y se trata de que sí hay un efecto extraño a los intereses del investigador o investigadora sea un efecto constante en todas las condiciones o grupos que tenga el estudio. La aleatorización supone asignar los tratamientos aleatoriamente a las unidades experimentales y de

esta forma las posibles causas comunes no afectan de forma sistemática a los resultados de la investigación, ya que se distribuyen de manera equilibrada. Es decir, no contaminan la relación entre las variables explicativas (variable independiente y variable dependiente) implicadas en la hipótesis de investigación. De este modo se elimina / controla el sesgo de selección (error sistemático o varianza sistemática secundaria), ya que los efectos de las posibles causas comunes se distribuyen de forma azarosa en las diferentes condiciones de tratamiento y se supone por lo tanto que su presencia está equilibrada o compensada en los diferentes grupos o condiciones de tratamiento (grupos equilibrados).

Solamente la metodología experimental permite plantear y comprobar hipótesis de causalidad propiamente dichas. En la metodología cuasi-experimental y no experimental el control del sesgo de selección requiere mantener constantes todas aquellas variables que teóricamente podrían atentar contra la validez de los resultados, pues diferenciarían a los sujetos y, por lo tanto, ocultarían se hubo o no un efecto del tratamiento o quizás lo confundirían.

En resumen, en la *metodología experimental* es el investigador o investigadora quién decide qué tipo de intervenciones va a evaluar en su estudio, provocando con ello la manipulación de la variable independiente con la intención de modificar la vida de los sujetos. Posteriormente, asigna aleatoriamente las condiciones de estudio creadas a los sujetos cuando se trata de un diseño entre-grupos (*diseños entre-grupos*) o, cuando los participantes reciben más de un tratamiento, aleatoriza el orden de presentación de los tratamientos cuando se trabaja con un *diseño intra-sujetos* o de medidas repetidas (ver Figura 10).

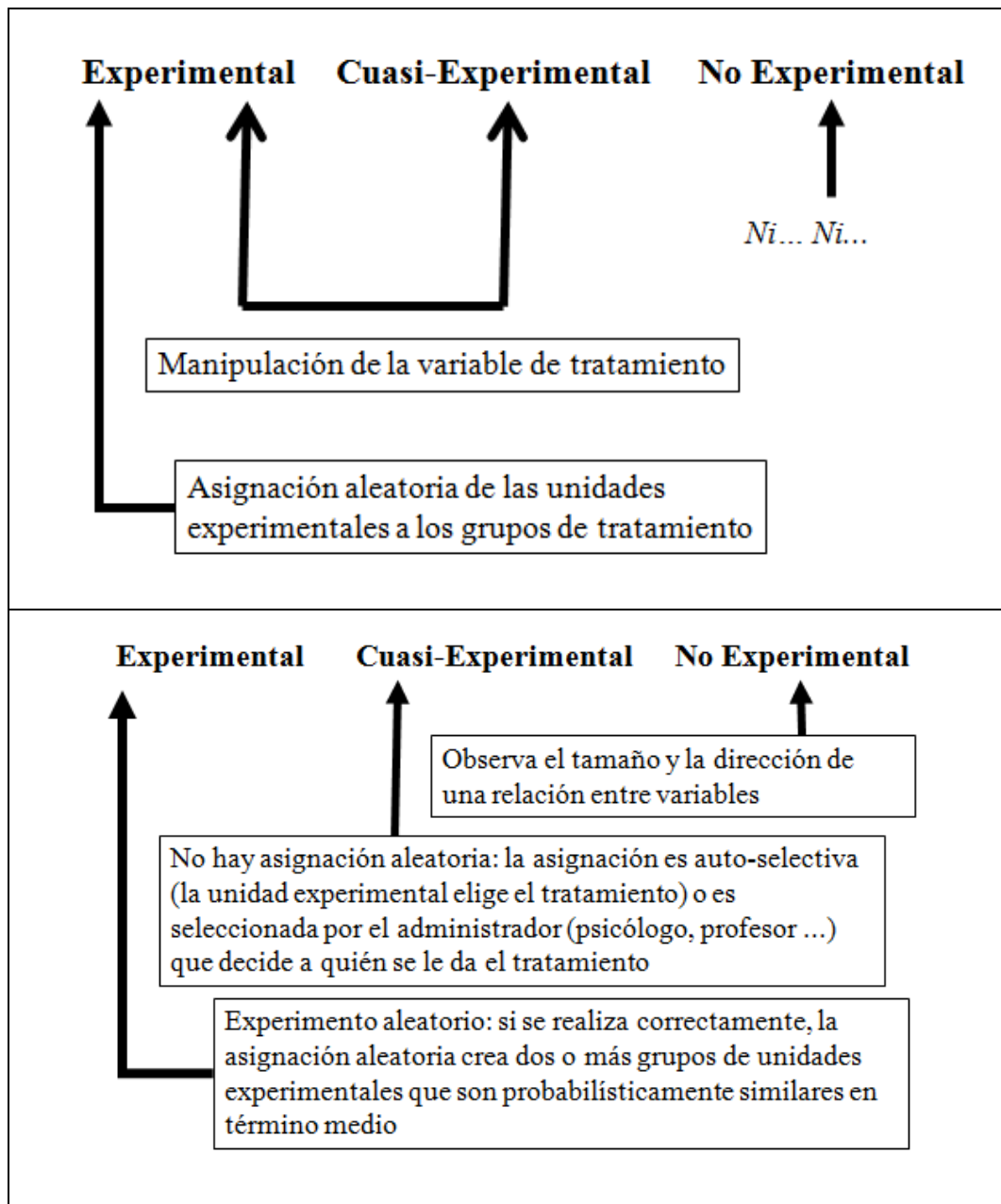


Figura 10. Características de las metodologías de investigación

En la metodología *cuasi-experimental* sí existe una manipulación de las causas o variables independientes, pero el tratamiento no se asigna de forma aleatoria a las unidades experimentales o sujetos o el orden de presentación de los tratamientos no es aleatorio. Se trabaja con grupos ya creados previamente, conocidos como ‘grupos intactos’ que no pueden formarse al azar. Aquí el control de las variables de las posibles variables de sesgo es menor que en la metodología experimental y requerirá técnicas de control específicas como la constancia (*diseños de bloques* por ejemplo), la eliminación de ciertas variables elaborando de

forma detallada determinados criterios de inclusión y exclusión de los participantes o del contexto de investigación o diseños más sofisticados como el *diseño con variables covariadas* o la utilización de *diseños con puntuaciones de propensión*.

En la metodología *no experimental* no existe manipulación de variables (por lo tanto no es posible la asignación aleatoria y como consecuencia tampoco la asignación aleatoria), solamente se observa el tamaño y la dirección de la relación encontrada entre las variables. No se produce ningún tipo de intervención o cambio deliberado sobre el sujeto, solamente se observa el constructo o variable tal y como la manifiesta el sujeto.

Otras clasificaciones de las metodologías de investigación

Dentro del área de la biomedicina, la clasificación más común de la metodología de investigación distingue entre estudios experimentales y estudios observacionales, diferenciándose por cómo se asigna el tratamiento (variable independiente).

En los estudios experimentales el investigador o investigadora asigna aleatoriamente las condiciones de tratamiento a los sujetos (conocido como ‘ensayo clínico aleatorio’ o aleatorizado, ECA).

En los estudios observacionales no existe asignación aleatoria de manera que los criterios de selección de los grupos de la población se basan en la presencia (casos) o no (controles) del desenlace objeto de investigación (resultado) en la muestra (diseños de casos y controles) o en la exposición al tratamiento sin asignación aleatoria (diseños de cohortes) (ver Figura 11).

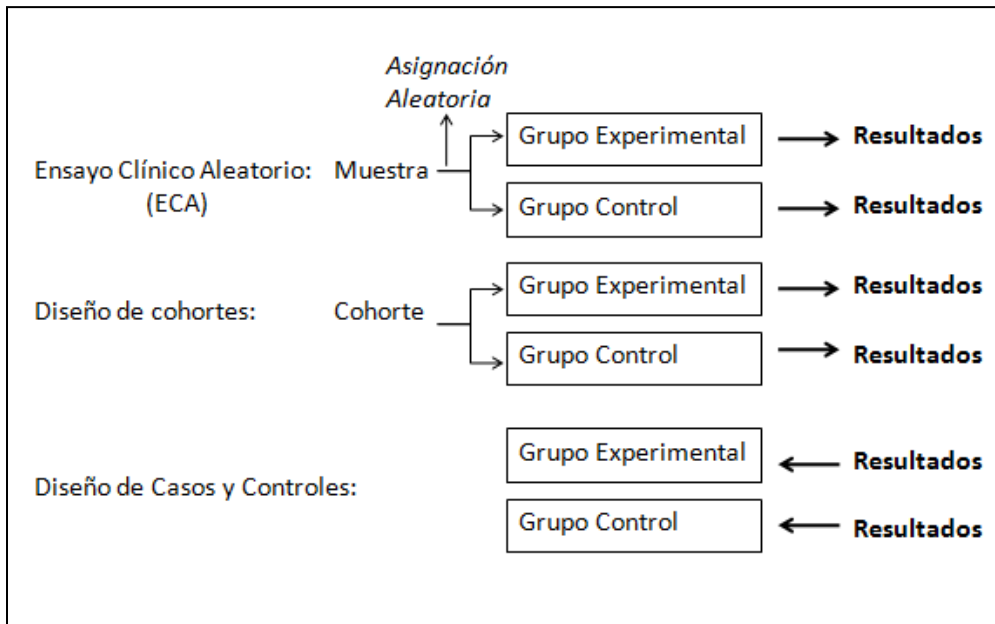


Figura 11. Configuración de los diseños ECA, cohortes y casos y controles

La principal diferencia entre el diseño ECA y un *diseño de cohortes* es la presencia o no de asignación aleatoria dado que en ambos casos el tratamiento antecede al efecto del tratamiento que se pretende evaluar. En los diseños de cohortes se mantiene la secuencia temporal de los acontecimientos tratamiento → resultado y suele requerir tamaños muestrales grandes que son seguidos durante largos períodos de tiempo para establecer conexiones entre las variables. El diseño de cohortes no es muy recomendable cuando se trata de analizar acontecimientos que se presentan con baja frecuencia porque resulta muy costoso. Se trata de seleccionar una cohorte de sujetos que han sido expuestos a la variable de interés para poder comparar sus puntuaciones con otra cohorte de sujetos que no han sido expuestos a dicha variable y medirlos, preferiblemente durante periodos largos de tiempo para comparar sus puntuaciones en la variable dependiente.

A diferencia de los diseños ECA y los de cohortes, en *el diseño de casos y controles* se parte del resultado, es decir, si los sujetos tienen ('casos') o no tienen ('controles') el resultado de interés, y se evalúa retrospectivamente la presencia o no de los posibles tratamientos o exposiciones. Este diseño es especialmente útil para estudiar resultados poco frecuentes. Esa inversión temporal en el desarrollo de los acontecimientos es una de sus principales debilidades a la hora de poder

establecer relaciones causales entre las variables de causa y efecto. Este diseño está limitado a una sola variable de resultado que es sobre la que se muestrea la población de casos y controles.

En resumen, se habla de ‘experimentos aleatorios’ o ‘ensayo clínico aleatorizado’ cuando hay manipulación de variables y asignación aleatoria y el ‘experimento aleatorio controlado’ señala que es un diseño aleatorio con un grupo de control para efectuar comparaciones entre el grupo experimental y el grupo de control. El clásico ensayo clínico o experimento aleatorio controlado se corresponde con la metodología experimental en las Ciencias Sociales y de la Salud y los diseños con un grupo de tratamiento y otro de comparación o control. En ambos casos es posible hablar de inferencia causal dado que hay control de la estrategia de asignación aleatoria de las diferentes condiciones de tratamiento.

Proceso de diseño de un estudio con metodología experimental

El diseño más sencillo de una investigación con una metodología experimental utilizando un *diseño entre-grupos unifactorial univariado* (diseño con una sola medición donde se trabaja con grupos de comparación formados por sujetos diferentes, con una sola variable independiente y una sola variable dependiente) incluye:

- 1) una variable independiente (A) con dos condiciones o grupos (a_1 y a_2)
- 2) una variable dependiente (Y)

Las unidades experimentales (generalmente se trata de sujetos) que quedan ubicadas dentro de cada condición o grupo deben ser todo lo similares que se pueda en todas las variables previas a la introducción o presentación del tratamiento o intervención. Con ello se trata de controlar las posibles variables contaminadoras o extrañas que afectan a la relación de las variables explicativas (independiente - dependiente) implicadas en la hipótesis de trabajo (amenazaría a la validez interna de los hallazgos).

En un diseño adecuadamente planificado, el objetivo es que exista homogeneidad entre las unidades experimentales que forman los grupos antes de

la introducción del tratamiento. Es decir, el objetivo es disponer de “grupos equilibrados o grupos homogéneos”, ya que deberían ser homogéneos en todas las variables antes de la introducción de la variable de tratamiento, dado que se formaron al azar (técnica de control de la aleatorización). Si así fuese entonces se podrá observar de forma correcta la magnitud del efecto o la magnitud de la relación entre las variables que forman la hipótesis del estudio.

Esa situación de investigación solo se produce cuando las condiciones de la investigación permitan que el azar actúe de forma adecuada, por ejemplo, con muestras grandes esa configuración de la homogeneidad es más probable. Sin embargo, a pesar de que se pueda disponer de unas condiciones óptimas de asignación aleatoria conviene tener presente que no es garantía absoluta de que los grupos sean equivalentes. Por ello, si el investigador conoce (por estudios anteriores o después de revisar la literatura) que una tercera variable o variable extraña puede ejercer un efecto contaminante sobre la relación entre las variables explicativas de la hipótesis de investigación (variable independiente y dependiente) entonces debe planificar su control directo en el diseño del estudio ya sea por ejemplo, manteniendo constante su efecto (diseño de bloques o apareamiento de dicha variable en los grupos), eliminando dicha variable, si es posible, del estudio (quitar esa condición, por ejemplo solo utilizar hombres) o quizás introduciéndola como una variable covariada que ajuste matemáticamente la relación entre la variable dependiente y la independiente (diseño de covarianza).

Para llevar a cabo la asignación aleatoria del tratamiento se puede utilizar, por ejemplo, la plataforma on-line ‘research randomizer’ (<https://www.randomizer.org/>) que es una herramienta gratuita que permite generar números aleatorios o asignar los participantes de forma aleatoria a las condiciones de la investigación.

En resumen, la técnica de control de la asignación aleatoria trata de homogeneizar a los grupos, ya que gracias al azar se supone que se distribuye aleatoriamente todas las variables (obrevadas y no observadas) que podrían crear diferencias entre los grupos o condiciones de la variable independiente y, solamente, el efecto de la variable independiente debería ser la causa de la diferencia entre ellos, si tal efecto existe.

Por supuesto, el éxito del procedimiento de la asignación aleatoria de la intervención está muy relacionado con el tamaño de la muestra, pues es necesario tener suficiente muestra para que actúe el azar. Y, además, el azar es caprichoso y podría no ser efectivo, ya que, a pesar de utilizar la aleatorización, podría dar lugar a grupos no homogéneos. De ahí, la importancia de considerar el control directo de la variable extraña o contaminadora por parte del investigador utilizando otras técnicas de control, tal y como ya se ha comentado, como la eliminación de la variable extraña si es posible, o factorizando su presencia en el diseño como un factor más en la ecuación estructural tal y como ocurre en los diseños de bloques no aleatorios o controlando su efecto matemáticamente tal y como se lleva a cabo en los diseños con variables covariadas.

Cuando se trabaja con una metodología cuasi-experimental (hay una intervención o una variable manipulada, pero no hay asignación aleatoria de las condiciones) también es necesario controlar que las unidades experimentales o sujetos que se van a comparar en cada grupo sean lo más homogéneas posibles en todas las variables (control del sesgo) para poder observar con validez las relaciones o la diferencia en la variable que es objeto de estudio en la investigación. También la metodología no experimental está expuesta al problema de la falta de control de las diferencias entre los sujetos de los grupos que se comparan más allá de la diferencia que es objeto de estudio en la investigación.

Especialmente en aquellas situaciones de investigación donde hay ausencia de aleatorización en la creación de los grupos es importante definir de forma clara y precisa un listado de variables de inclusión y exclusión dirigido a construir grupos o condiciones donde se hayan controlado en la fase de planificación aquellas variables que la literatura ha destacado como relevantes y que afectarían al estudio de la relación entre la variable independiente y dependiente que forman la hipótesis de investigación del estudio. Una vez ejecutados los criterios de inclusión y exclusión, el investigador o investigadora podría proceder con la recogida de los datos de su investigación. Esos criterios deben justificarse teóricamente y aplicarse siempre a priori, es decir, antes de recoger los datos. Y, por supuesto, deberán ser explicados en el informe de investigación con argumentos.

Por lo tanto, una vez se ha seleccionado una muestra de participantes (si el muestreo es probabilístico mucho mejor, ya que actúa la técnica de control de la aleatorización y se proporciona validez externa a los resultados), si es posible, se introduce la asignación aleatoria del tratamiento a dichas unidades experimentales o sujetos (las condiciones de la denominada variable de tratamiento o variable independiente A se asignan al azar a cada sujeto). Es decir, se trabajaría con una metodología experimental donde un grupo de sujetos recibe el tratamiento objeto de estudio (a_1 grupo experimental) y el otro grupo recibe otro tratamiento (a_2 grupo de control o grupo de comparación) que puede ser un tratamiento de comparación, un tratamiento de placebo, un tratamiento de lista de espera, otro tipo de tratamiento... Al final del estudio y si los grupos están balanceados en todas las variables previas al tratamiento (grupos previamente homogéneos o equivalentes, es decir, se han controlado las variables extrañas), cualquier diferencia entre las puntuaciones (variable dependiente Y) de los grupos posterior a la introducción del tratamiento se asume que estará causada por el efecto del tratamiento. Se trataría en este caso de un diseño con una metodología experimental, ya que tiene 1) manipulación de la variable independiente y 2) asignación aleatoria del tratamiento a los sujetos y, solamente con este tipo de metodología se podrán realizar interpretaciones causales entre las variables, es decir, interpretar que la variable independiente (manipulada y asignada al azar) es la causa del efecto detectado en la variable dependiente medida.

A continuación se detallan dos ejemplos de investigación donde se trabaja con una metodología experimental.

Ejemplo 1. En el año 2006 Dar-Nimrod y Heine realizaron un experimento para tratar de analizar si la exposición a teorías científicas afectaba al rendimiento de las mujeres en matemáticas, publicando sus hallazgos en la revista *Science*. En concreto, analizaron si tener la creencia de baja aptitud para las matemáticas de las mujeres afectaba a su rendimiento.

La hipótesis sustantiva mantiene que dicha creencia se convierte en realidad cuando la mujer cree que lo es. Los investigadores crearon tres grupos (A = 3, variable independiente o factor A con tres condiciones: a_1 , a_2 , y a_3) y asignaron de

forma aleatoria a las mujeres a uno de los grupos (*randomization*, R, o aleatorización).

La variable que se manipuló (variable independiente A) fue la creencia sobre el origen de las diferencias entre sexos en el rendimiento en matemáticas (a través de la lectura de un texto): grupo a_1 origen genético, grupo a_2 origen basado en la experiencia o aprendizaje y grupo a_3 ideas estereotipadas generales.

Posteriormente se midió el rendimiento (porcentaje de respuestas correctas) en una prueba de matemáticas (variable dependiente Y).

El diseño tuvo un pre-test y un post-test. Se trata de un *diseño pre-test/post-test con grupo de control equivalente*, ya que los tratamientos (las condiciones de la variable independiente) fueron asignados al azar y de ahí que se utilice el término equivalente para describir al grupo de control. Si no hay asignación aleatoria del tratamiento (metodología cuasi-experimental) se rotularía como un *diseño pre-test/post-test con grupo de control no equivalente*, ya que al decir grupo de control no equivalente se señala que no hubo asignación aleatoria en la distribución del tratamiento. El diseño de la investigación del ejemplo 1 se representa gráficamente en la Figura 12.

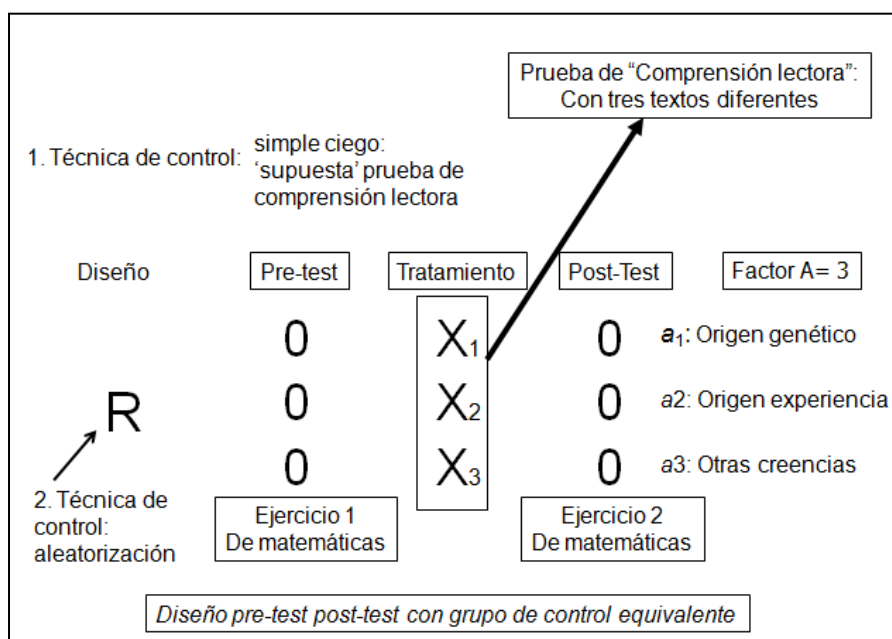


Figura 12. Configuración del diseño pre-test/post-test con grupo de control equivalente

Los resultados del estudio de Dar-Nimrod y Heine (2006) señalan que las mujeres sacan peor nota si creen sufrir una dificultad innata para las matemáticas.

Por el contrario, su rendimiento mejora cuando creen que las diferencias entre los sexos está provocada por la experiencia o el aprendizaje.

Como se ha comentado anteriormente, si la asignación hubiese sido no azarosa (metodología cuasi-experimental) entonces el diseño se formularía como *diseño pre-test/post-test con grupo de control no equivalente (non randomization, nR, no aleatorizado)*. Es decir, se trata de una metodología donde hay manipulación de la variable independiente, pero no se produce la asignación aleatoria del tratamiento a los sujetos para formar los grupos de comparación.

En la investigación de Dar-Nimrod y Heine (2006), junto con la técnica de control de la asignación aleatoria, se utilizó la técnica de control denominada “simple ciego”, es decir, los sujetos o participantes del estudio desconocían el verdadero objeto de análisis del estudio que eran sus creencias sobre el origen de las diferencias en aptitud matemática entre los hombres y las mujeres y se les indicaba que se trataba de una prueba para medir la comprensión lectora.

En ocasiones, se pueden utilizar otras técnicas de control de cegamiento como el ‘doble ciego’ donde ni el sujeto ni el investigador (por ejemplo) conocen qué tratamiento reciben los sujetos o el ‘triple ciego’ donde el sujeto, el investigador y el evaluador (por ejemplo) no conocen el tratamiento que recibe cada participante. En el informe de investigación es necesario detallar si se aplicó la técnica de control del cegamiento, quienes fueron cegados y cómo se llevó a cabo dicha técnica.

Ejemplo 2. En los estudio de Pavlov sobre el condicionamiento clásico se demostró que los perros tenían respuestas condicionadas ante un estímulo neutro (por ejemplo el plato de comida, un ruido o incluso los pasos del encargado al acercarse a los perros para ofrecer la comida) que había sido repetidamente asociado a un estímulo incondicionado como la comida. La respuesta condicionada de los perros ante el plato vacío de comida, ante el ruido de las llaves del encargado o al oír los pasos del encargado, hacía que los perros segregasen saliva sin la presencia del estímulo incondicionado de la comida.

Para mejorar el efecto del condicionamiento los resultados de la literatura recomiendan que el estímulo condicionado y el estímulo incondicionado sean de la misma modalidad sensorial. La modalidad sensorial común podría ser por ejemplo

el sentido de la vista tanto para el estímulo condicionado como para el estímulo incondicionado o podría ser el sentido del oído para captar ambos estímulos.

Supongamos que un investigador desea replicar los hallazgos de Pavlov y planifica un estudio sobre la segregación salivar en perros. Elabora dos situaciones experimentales (tratamiento o variable independiente $A = 2$) con 10 perros en cada condición ($n_1 = n_2 = 10$; por lo tanto $N = 20$; 'diseño ortogonal', ya que hay el mismo número de obseraciones en cada condición o grupo; si no hubiese el mismo número de obseraciones se llamaría 'diseño no ortogonal'). En una de las situaciones asocia el estímulo incondicionado de visionado de la comida con una luz intermitente (condición del tratamiento a_1). En la otra situación vincula el estímulo incondicionado de visionado de la comida con un ruido intermitente (condición del tratamiento a_2). Posteriormente, asigna aleatoriamente el tratamiento a_1 a diez animales y el tratamiento a_2 a otros diez animales (metodología experimental; diseño entre-grupos).

Teniendo en cuenta el conocimiento teórico previo que existe sobre la relación entre la modalidad sensorial y el efecto del condicionamiento clásico (mejora si el estímulo condicionado e incondicionado son de la misma modalidad sensorial) se plantea la siguiente hipótesis de investigación, hipótesis científica o teórica o hipótesis sustantiva: los perros que reciben el tratamiento a_1 (misma modalidad sensorial del estímulo incondicionado y el estímulo condicionado, (visual-visual) mostrarán la respuesta condicionada más rápidamente que el grupo que recibe el tratamiento a_2 donde los estímulos son de diferente modalidad sensorial (visual-auditiva).

En este diseño de investigación del estudio de Pavlov con dos grupos de tratamiento la variable independiente manipulada es el tipo de estímulos apareados (A) con dos condiciones (a_1 misma modalidad sensorial y a_2 distinta modalidad sensorial) y la variable medida o variable dependiente (Y) es el tiempo que tarda el perro en manifestar la respuesta condicionada ante el estímulo condicionado. El procedimiento de asignación aleatoria de las condiciones de tratamientos a_1 y a_2 a los perros (todos los perros tenían la misma probabilidad de recibir el tratamiento a_1 que el a_2 , es decir, un 50%) es la técnica que controla las posibles diferencias previas entre los animales que podrían afectar a los resultados de forma

sistemática y es propia de la metodología experimental. La técnica de control de la aleatorización distribuye de forma equilibrada todas aquellas posibles variables que podrían diferenciar a los animales de forma sistemática antes de la introducción del tratamiento o condición de intervención, creando ‘grupos equivalentes’ u homogéneos. De este modo, la comparación de las puntuaciones que se realizará después de las intervenciones de condicionamiento entre los animales de la condición a_1 y la de los animales en a_2 permite estimar el efecto de ambas intervenciones y si hubiese una diferencia estadísticamente significativa en el tiempo que tarda en manifestarse el condicionamiento (variable dependiente, Y) se podría concluir que ello es debido a la vinculación entre modalidades sensoriales de la misma naturaleza o de distinta naturaleza (variable independiente A).

Los argumentos que apoyan que la metodología experimental ofrece interpretaciones de causalidad se basan en la presencia de la aleatorización en la creación de los grupos, el control del sesgo y la manipulación de variables. Así, dadas las características de asignación aleatoria y control de la influencia de posibles variables extrañas –se trabaja con grupos previamente homogéneos– cualquier diferencia que se produzca en los grupos después de realizar el experimento podrá ser atribuida a la manipulación experimental (Alvira, Avia, Calvo y Morales, 1980).

En otras palabras, para poder hablar de relaciones causales entre las variables, la asignación aleatoria de los sujetos a los grupos experimentales es un elemento clave dado que permite asumir que las condiciones experimentales creadas o manipuladas no están relacionadas, excepto por azar, con otras causas alternativas. Y la metodología experimental incluye por definición la asignación aleatoria, la manipulación de variables y el control de sesgos por aleatorización. Fisher (1935) introdujo la asignación aleatoria en los diseños experimentales para asegurar que la única diferencia entre los grupos era la variable de tratamiento o variable independiente –eran grupos equivalentes estadísticamente hablando antes de la manipulación–, y por lo tanto si aparecían diferencias, su explicación vendría dada por su presencia.

En resumen, cuando se dice que un estudio se ha realizado con una metodología experimental se quiere señalar que se han creado, manipulado o

controlado diferentes valores o condiciones de la variable independiente, a cuyos niveles las unidades experimentales (generalmente individuos humanos) son asignadas aleatoriamente, con objeto de observar si se producen diferencias en la variable dependiente medida. Además, los diseños experimentales son los únicos que permiten realizar interpretaciones de causalidad donde la variable independiente es la causa del efecto observado en la variable dependiente.

Metodología cuasi-experimental

Establecer relaciones causa-efecto es muy difícil cuando no es posible asignar aleatoriamente las diferentes condiciones del tratamiento a los sujetos. En esta situación determinar el efecto de un tratamiento es problemático dadas las diferencias existentes entre el grupo experimental y el grupo de control en la fase de línea base o antes de la administración del tratamiento. En estos casos se trabaja con la denominada metodología cuasi-experimental donde sí existe manipulación de la variable independiente (está sometida al control del investigador o investigadora), pero no es posible llevar a cabo la técnica de la asignación aleatoria (se trata de un diseño con grupo de control no equivalente). Se trabaja con grupos ya creados previamente, conocidos como 'grupos intactos' y por lo tanto con posibles causas comunes que se desconocen o no están controladas. El control estadístico de las variables contaminadoras (diseño con variables covariadas) o la introducción de las variables en el modelo de diseño de investigación (diseño de bloques) pueden ayudar a lograr una cuasi-aleatorización y con ello plantear hipótesis de causalidad. Esto supone trabajar con diseños más complejos y con una técnica estadística más sofisticada que de alguna manera atrape las diferencias previas entre los grupos, controlando el sesgo de selección; por ejemplo, con diseños basados en la puntuación de propensión, diseños de bloques o análisis de la covarianza. Es decir, la falta de asignación aleatoria trata de ser compensada con una mayor sofisticación del diseño de investigación.

En opinión de Kirk (1995), el método cuasi-experimental se utiliza cuando no es posible la asignación aleatoria o cuando por razones prácticas o éticas es necesario utilizar grupos naturales o grupos ya formados como por ejemplo sujetos con una determinada enfermedad, sujetos que han sido sometidos a abuso sexual o sujetos maltratadores que reciben tratamiento de forma voluntaria.

En resumen, en la metodología experimental y en la cuasi-experimental hay una intervención bajo el control del investigador dado que decide qué tipo de tratamientos va a evaluar en su experimento (manipulación de la variable independiente), pero sólo en la metodología experimental hay asignación aleatoria de las condiciones de la variable independiente a los sujetos o grupos. Qué supone esto. Pues que en la metodología cuasi-experimental la posible presencia de diferencias previas entre los grupos (sesgo de selección) es una fuente de sesgo que será necesario controlar mediante la planificación del diseño y la introducción de las posibles causas comunes en el modelo de investigación (por ejemplo con el diseño de bloques) o con el uso de herramientas estadísticas como por ejemplo el diseño de covarianza.

Metodología no experimental

En la metodología no experimental se encuentran los denominados estudios observacionales y los estudios de encuesta. En ambos casos no existe ni manipulación de la variable independiente -los valores de la variable independiente son sólo observados—, ni por supuesto asignación aleatoria de las condiciones de la variable independiente a los sujetos. Se trata de estudiar los fenómenos tal y como ocurren de forma natural (Anguera, 2010). El planteamiento de las relaciones entre las variables es de covariación y nunca causal. Como señalan Shadish y Cook (2002), un principio muy conocido en la investigación es que “la correlación no prueba la causación”. Es decir, si no se conoce qué ocurre en primer lugar (antecedente vinculado a la variable independiente) entonces las posibles explicaciones alternativas del desenlace (o consecuencia que se observa en la variable dependiente) se multiplican.

En los diseños no experimentales (en ocasiones referidos como diseños correlacionales aunque el término conduce a la confusión entre la técnica estadística y la metodología de investigación) se plantea la estimación de efectos, pero su estructura no permite identificar la causa y el efecto. No hay variables manipuladas ni tampoco efectos identificados como antecedente y consecuente. No hay asignación aleatoria. Su ejecución sólo permite conocer la magnitud y la dirección de los efectos estimados con el diseño de la investigación.

En los estudios llevados a cabo con una metodología observacional, la manipulación de los grupos es una característica propia de los individuos cuya naturaleza es precisamente el objetivo de investigación del estudio, configurándose entonces el grupo de tratados como grupo de ‘expuestos’ a la variable de interés y el grupo de no tratados como grupo de ‘no expuestos’ a dicha variable. Ahora ya no es el investigador o investigadora la persona que decide qué tipo de condiciones tiene el factor o la variable independiente pero sí es la persona que decide qué niveles de la variable independiente va a estudiar. Es decir, qué grupos de expuestos o de no expuestos a las condiciones de la variable independiente va a analizar en su trabajo. Por lo tanto, ahora los individuos ya tienen la característica que define al grupo y es justamente esa característica (sus diferencias) la que define el objetivo de la investigación. Por ejemplo, sujetos que toman alcohol frente a sujetos que no ingieren alcohol, sujetos fumadores o no fumadores, sujetos que reciben cursos para búsqueda de empleo o sujetos que no reciben el curso, una mujer o un hombre que trabajan fuera del hogar o están desempleados, madre o padre fumadora o madre o padre no fumadora, familia nuclear o familia con padres divorciados.

Por lo tanto, en los estudios observacionales (no experimentales) no existe una manipulación directa de la variable de tratamiento tal y como sí sucede en los diseños con metodología experimental y cuasi-experimental. Se trata de ‘diseños prospectivos’, ya que los sucesos o resultados han ocurrido antes de que el estudio comenzara. Por ejemplo, el sujeto ya es fumador o ya está desempleado. Por ello, la metodología no experimental se aplica en aquellos casos donde el investigador o investigadora no puede presentar los niveles de la variable independiente a voluntad propia ni puede crear los grupos experimentales por aleatorización, aunque sí puede introducir algo similar al diseño experimental en su programación de procedimientos para la recopilación de datos como el cuándo y el a quién de la medición (Campbell y Stanley, 1963). Clásicos ejemplos de estudios observacionales son los de la salud y el tabaco o la ingesta de vitamina C y la prevención del cáncer.

Como ya se ha comentado, la gran ventaja de los experimentos aleatorios (metodología experimental) es la posibilidad que ofrecen para estimar de forma no sesgada los efectos causales dado que la asignación aleatoria del tratamiento

permite que el grupo control y el grupo experimental sean iguales en términos medios en todas las características, tanto observadas como no observadas, siendo la variable de tratamiento la única diferencia entre ellos. Por supuesto, si la técnica de la asignación aleatoria funcionó de forma correcta.

Tanto en la metodología cuasi-experimental como en los estudios observacionales o no experimentales el investigador no controla la asignación de los tratamientos a los sujetos participantes y por ello no puede estar seguro de si sujetos similares reciben tratamientos diferentes, es decir, si son grupos comparables entre sí. En estos casos el denominado ‘sesgo de selección’ contamina los resultados y por lo tanto amenaza a la validez interna de los hallazgos.

En los estudios de encuesta no existe intervención o tratamiento y por lo tanto no cabe preguntarse por el control de la asignación aleatoria. Se trata de estudios de opinión donde sí existe un control de la selección de la muestra si se aplica el muestreo probabilístico. Ejemplos de encuesta son las realizadas para conocer la intención de voto y poder con ello predecir los resultados de las elecciones, predecir el uso de productos o medir los cambios de opinión. Las cuestiones vinculadas con el muestreo aleatorio (probabilístico) son la clave metodológica que garantiza la validez de los resultados obtenidos. Definir la población objeto de estudio, determinar el tamaño de la muestra y el método de extracción de los elementos de dicha muestra y fijar el error muestral asumible en la investigación son las decisiones más importantes que el investigador debe adoptar ante el diseño de un estudio con metodología de encuesta o selectiva.

Asignación aleatoria del tratamiento

La asignación aleatoria del tratamiento a cada uno de los sujetos (unidades experimentales) que forman la muestra de participantes tiene como objetivo garantizar que la posible función causal hallada al final del estudio no está provocada por otras causas (denominadas ‘terceras variables’ o variables extrañas) ajenas al efecto de la variable independiente.

La ventaja de la técnica de la aleatorización es que trata de prevenir diferencias sistemáticas entre los grupos de cualquier tipo ya estén identificadas por el

investigador (conocidas por el investigador, denominadas variables covariadas observadas), o no identificadas (variables covariadas no observadas).

La asignación aleatoria implica que cada unidad experimental o participante tiene la misma probabilidad de formar parte de un grupo u otro. Por ejemplo, si el diseño tiene dos grupos ($A = 2$) entonces cada sujeto tiene una probabilidad de $\frac{1}{2}$ de pertenecer al grupo a_1 y la misma probabilidad ($\frac{1}{2}$) de pertenecer al grupo a_2 (ver Figura 13). Gracias a la técnica de la aleatorización disminuye la probabilidad de diferencias sistemáticas previas al tratamiento entre los grupos experimental y control, es decir, hay un control del error sistemático o sesgo.



Figura 13. Proceso de asignación aleatoria de los sujetos (randomización)

En ocasiones puede ocurrir que la asignación aleatoria se produzca después de estratificar a los sujetos en alguna variable no aleatoria (por ejemplo en un diseño con variables bloqueadas donde la variable bloqueada es el factor de bloqueo) que se desea mantener controlada por constancia, configurando diseños 'parcialmente aleatorios' si la variable independiente o factor del diseño vinculado a la varianza sistemática primaria o del efecto se asigna al azar.

Por ejemplo, los sujetos pueden clasificarse previamente a la introducción del tratamiento en una determinada variable como el nivel de glucosa en sangre en la línea base: alta, media y baja, y posteriormente se procede a la asignación aleatoria del tratamiento a los grupos experimental y control dentro de cada bloque. Es decir, con ello se controla que en todos los grupos de tratamiento hayan sujetos con los tres bloques de glucosa previa.

Así, los sujetos que forman el bloque de nivel alto de glucosa son asignados al azar a cada una de las dos condiciones de tratamiento, del mismo modo los del grupo de glucosa media y finalmente los del grupo de glucosa baja (ver Figura 14).

De este modo se garantiza la distribución homogénea de los distintos niveles de glucosa en el grupo experimental y en el grupo de control. Se trata en este caso de un *diseño de bloques con restricciones en la aleatorización* porque no todas las variables o factores del diseño se han asignado al azar (la glucosa previa en sangre es una variable asignada y por lo tanto no es posible aleatorizarla).

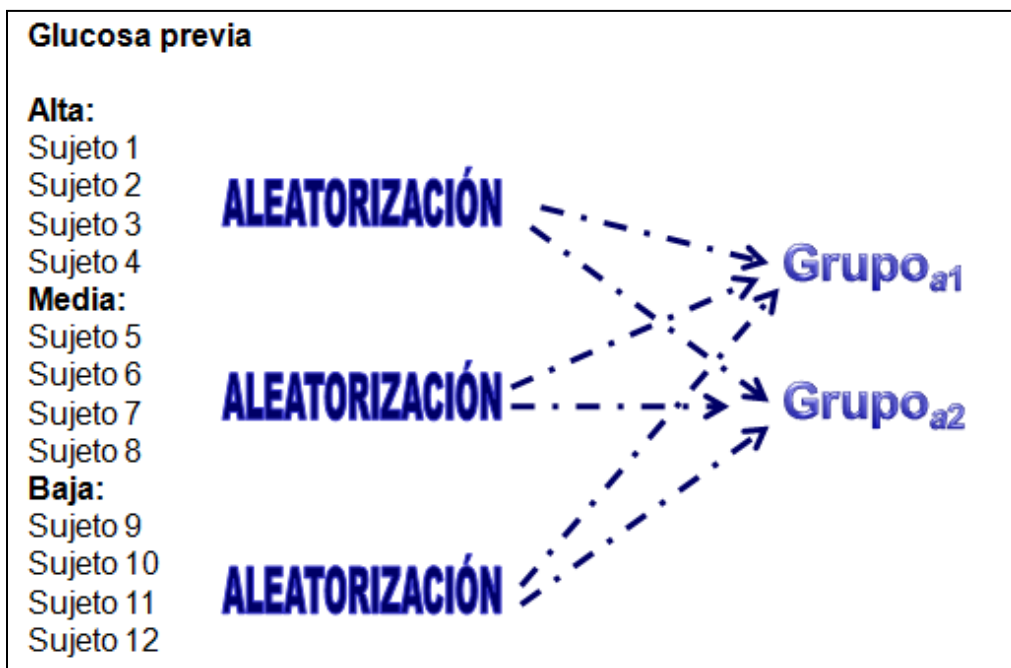


Figura 14. Bloqueo de la variable nivel de glucosa previa

Y el *diseño de bloques con restricciones en la aleatorización* tendrá, al menos, dos factores: factor A vinculado a la variable independiente Grupo (factor de 'varianza sistemática primaria') y el Factor B vinculado a la variable de bloqueo que se controla (factor de 'varianza sistemática secundaria' controlada), configurando un *diseño entre grupos de bloques A x B parcialmente aleatorio*.

Evidentemente, no hay dos grupos que sean exactamente iguales antes del tratamiento. Incluso aunque se utilice la técnica de la asignación aleatoria pueden existir diferencias individuales o problemas relacionados con el muestreo (errores

de muestreo) que provocan diferencias entre los grupos. Se trata del error muestral (varianza no sistemática).

La técnica estadística determina la probabilidad de que las diferencias observadas entre los grupos podrían ser debidas al mismo proceso de asignación aleatoria de las unidades experimentales que forman los grupos. Si la probabilidad (valor p de probabilidad) del estadístico utilizado en la prueba del contraste de hipótesis del estudio es baja (generalmente $p \leq .05$) entonces se asume que las diferencias entre los grupos están causadas por el tratamiento interpretándose como 'diferencias estadísticamente significativas'.

Por el contrario, si la probabilidad de que las diferencias encontradas entre los grupos debidas al error de muestreo es alta ($p > .05$) entonces se asume que las diferencias observadas entre los grupos son debidas al mismo proceso de asignación aleatoria y se concluye que las diferencias encontradas entre las puntuaciones de ambos grupos 'no son estadísticamente significativas'.

Es decir, asumiendo que las grandes o las pequeñas diferencias entre las puntuaciones de las condiciones son debidas al azar o al error de muestreo (planteamiento de la hipótesis nula), se obtiene el valor p de probabilidad de los datos o resultado del estudio que se ha realizado y si es un valor pequeño se puede concluir que la variabilidad detectada se puede atribuir a algo más que al error de muestreo o al azar, concluyendo con el rechazo de la hipótesis nula que conduce, como consecuencia, a la aceptación de otra hipótesis llamada hipótesis alternativa. Posteriormente se explicarán con detalle las hipótesis estadísticas de hipótesis nula e hipótesis alternativa.

Por lo tanto, disponer de grupos lo más homogéneos posibles en todas las variables antes de la introducción del tratamiento o variable independiente es fundamental para abordar con calidad los resultados aportados por el estudio (validez de los resultados). Y es la metodología de investigación del estudio la que determina la jerarquización de las pruebas o la evidencia aportada por los estudios basándose en la técnica empleada para controlar las variables y para formar grupos equilibrados o balanceados, es decir, grupos equivalentes.

En resumen, se habla de 'experimentos aleatorios', 'ensayo clínico aleatorizado', 'experimentos aleatorios controlados' o 'metodología experimental'

cuando hay manipulación de variables y asignación aleatoria del tratamiento (mayor control de la validez interna). El clásico ensayo clínico o experimento aleatorio controlado se corresponde con la metodología experimental en las Ciencias Sociales y de la Salud. En ambos casos es posible hablar de inferencia causal dado que hay control de la estrategia de asignación aleatoria de las diferentes condiciones de tratamiento y, por lo tanto, se asume que las diferencias entre los grupos o condiciones se pueden atribuir al efecto de la variable independiente manipulada.

Manipulación y aleatorización y, por supuesto, el mayor grado de control de la situación que implican ambos elementos, resume las características más sobresalientes de la estrategia experimental.

El concepto de manipulación de una variable supone una intervención deliberada o el control de los valores que se seleccionan y administran. Administrar un tratamiento equivale a definir un valor de la variable independiente y en el contexto del diseño de investigación no tiene connotación ética alguna. El elemento de la aleatorización permite controlar fuentes de sesgo y asegura la aplicación correcta de las técnicas estadísticas del contraste de hipótesis.

Si la aleatorización es completa se produce tanto en la selección de las unidades experimentales (relacionada con la validez externa de los resultados) como en la asignación a las condiciones de la variable independiente (relacionada con la validez interna de los resultados). De este modo, gracias a la 'selección aleatoria' de la muestra, el diseño permite trabajar con muestras representativas cuya información refleja estimaciones exactas de los parámetros que describen estadísticamente a la población (Arnau, 1989a). Las pruebas estadísticas están basadas en la asunción del muestreo aleatorio a partir de la población, siendo esencial el muestreo aleatorio o la asignación aleatoria de las unidades observacionales a los grupos de tratamiento para la aplicación correcta de dichas pruebas, asegurando la independencia de las observaciones o de los errores así como la estimación correcta de los parámetros. En otras palabras, una prueba de significación estadística utilizada sin aleatorización no facilitará información válida sobre la probabilidad de un resultado bajo el modelo de la hipótesis nula (Shaver, 1993).

Diseño de $N = 1$

Existen otras alternativas metodológicas, como por ejemplo, el diseño de caso único o diseño de $N = 1$, que resulta difícil de incluir en cualesquiera de las distintas opciones de clasificación de los métodos de investigación que se han señalado anteriormente. Este diseño surgió dentro de la orientación experimentalista propia del conductismo, y creemos que por sus peculiaridades representa un método de investigación propio que es especialmente útil y muy utilizado dentro del ámbito de la psicología aplicada y en especial en la investigación clínica.

Un criterio amplio de clasificación de los diseños experimentales distingue entre diseño experimental clásico o fisheriano (diseño de $N > 1$) y diseño de $N = 1$. Cook y Campbell (1986) diferencian dos modelos de diseño experimental en función de la naturaleza del control que aplican, distinguiendo entre la tradición del control estadístico, propio de los diseños de comparación de grupos o $N > 1$, y la tradición del control y aislamiento experimental que caracteriza a los diseños de replicación intrasujeto o $N = 1$.

El diseño de $N = 1$ o diseño conductual ha recibido diversos rótulos; como señala Arnau (1995a), el diseño ha sido etiquetado como diseño de replicación intra-sujeto (Arnau, 1984; Gentile, Roden y Klein, 1972), especialmente dentro del ámbito clínico, como, diseño de caso único (Barlow y Hersen, 1984; Kazdin, 2010), diseño operante (Sidman, 1973) o diseño de un sólo sujeto (Bordens y Abbot, 1988; Cozby, 1993). Nosotros nos hemos quedado con el título menos comprometedor de diseños de $N = 1$ (Pascual-Llobell, Frías-Navarro, y García-Pérez, 1995).

En el análisis histórico que realizan Barlow y Hersen, (1984), señalan que la investigación experimental se desarrolló en dos direcciones:

- las investigaciones sobre grupos diferentes donde se estudia el grupo y su respuesta media como reflejo de la acción de la variable de tratamiento, actuando la aleatorización como principio fundamental de homogeneización de los grupos y
- las investigaciones sobre una sola unidad de observación (sujeto o grupo reducido) que es sometida a los distintos tratamientos en una

situación de estricto control experimental donde la fase de línea base proporciona la puntuación de comparación y control de variables.

En estas últimas investigaciones, la replicación de los datos se basa en la reversibilidad de los tratamientos y, a diferencia de las investigaciones de grupos, la generalización de los resultados a la población de origen se hace difícil. Sin embargo, también es cierto que el objetivo prioritario de un experimento verdadero no es realizar estimaciones probabilísticas del efecto en la población (Berkowitz y Donnerstein, 1982) sino que se centra en inferir relaciones entre variables, eliminando posibles causas alternativas, posibilitando con ello la creación de teorías.

Desde una perspectiva vinculada con la tradición social aplicada y en conexión directa con la incorporación de la dimensión temporal que intrínsecamente el propio diseño incluye, los diseños de $N = 1$ son conocidos como diseños de series temporales (Arnau, 1995b) rotulados como diseño de series temporales interrumpidas por Campbell y Stanley (1966), quienes lo sitúan como método cuasi-experimental dada la ausencia de aleatorización, o identificados simplemente como diseños de series temporales (Glass, Willson y Gottman, 1975; Hayes, 1981; Kratochwill, 1978).

La ubicación de la naturaleza de la metodología de caso único no es unánime entre los investigadores y hay quienes lo sitúan en el ámbito cuasi-experimental y otros en el propiamente experimental. Por supuesto, existen ejemplos de aplicación de diseño de $N = 1$ que son perfectamente experimentales del mismo modo que también hay trabajos cuya metodología es cuasi-experimental (Cook y Campbell, 1979; Edgington, 1987).

Siguiendo la proposición de los autores que sitúan este diseño dentro del paradigma experimental (por ejemplo Arnau, 1995b; Barlow y Hersen, 1984; Kazdin, 1992), el diseño de $N = 1$ reúne las condiciones de método experimental de investigación siempre que su planteamiento incluya las condiciones de control estricto de los efectos nocivos provocados por variables extrañas, donde la fase de línea base tiene un protagonismo destacado como elemento de comparación. Además, y siguiendo los comentarios de Arnau (1995b, p. 177), el carácter experimental de los diseños de $N = 1$ se expresa en los siguientes términos:

“... se trata de estructuras donde no sólo tiene que estar presente el control de las posibles fuentes de confusión, sino la manipulación de la variable de tratamiento así como la correcta especificación de la variable de medida o de resultado. Comparte, pues con el diseño experimental clásico un objetivo común: evaluar la acción causal de la variable independiente y establecer el grado de impacto o efecto que ejerce dicha variable sobre alguna medida de respuesta del sujeto. Quizá podríamos destacar, como característica propia de esta estructura, el hecho que se suele utilizar como unidad de análisis a un solo individuo o a un reducido grupo de individuos, con la consiguiente repercusión que ello tiene en la generalización de resultados”.

Los problemas de validez externa pueden ser obvios desde el momento que se utiliza un sólo sujeto (o pocos sujetos) y la generalización de los datos puede ser ampliamente cuestionada. En los diseños de $N = 1$ la replicación es la clave de la generalización y la validez interna del diseño.

La replicación implica generar nuevos estudios que especifican claramente las condiciones de tratamiento y medida que se desean replicar. Barlow y Hersen (1984) describieron tres tipos distintos de replicación: *directa* donde se utilizan los mismos tratamientos con nuevos sujetos, *sistemática* donde hay un cambio de las variables de interés como contexto, tipo de desajuste o trastorno... y *clínica* donde se comprueba el tratamiento con sujetos que presentan problemas conductuales semejantes. Es evidente, que el primer tipo de replicación tiene que ver con la fiabilidad de la investigación y, por tanto, se relaciona principalmente con la validez interna de la misma. En cambio, los otros dos tipos de replicación tienen que ver con la validez externa. Los autores afirman que “... una serie de diseños de caso único con sujetos semejantes a los que se les aplica el mismo tratamiento original - replicación-, puede sobrepasar en tres o cuatro veces el diseño experimental con grupos de tratamiento / grupo de control” (Barlow y Hersen, 1984, p. 57).

Otras amenazas contra la validez interna, por ejemplo, la interacción selección por tratamiento, son difíciles de controlar en este tipo de diseños: un sujeto particular con unas características individuales, puede reaccionar de forma específica al tratamiento, haciendo igualmente difícil la atribución causal que se pretende.

En estos diseños de $N = 1$, al igual que en los diseños cuasi-experimentales, hay que extremar las condiciones de control y tomar las precauciones necesarias para que factores tales como la historia, la maduración... no afecten seriamente a la validez de los resultados.

Otro tipo de amenaza contra la validez interna, y que puede afectar también a la validez de conclusión estadística, procede del hecho mismo de la repetición de la secuencia tratamiento - no tratamiento. Cuando a un sujeto se le dan múltiples tratamientos de manera serial o el mismo tratamiento en múltiples etapas, ¿hay interferencia? ¿Hay facilitación? ¿Los efectos de un tratamiento dependen del orden en que se hayan administrado? ¿Es diferente el efecto según sea en presencia o ausencia de un segundo elemento? De nuevo la salida airosa que, como solución general, se puede aportar parece venir de la replicación de secuencias alternativas.

Por otro lado, también las estrategias de reversibilidad del tratamiento, cuando es éticamente posible, son otro elemento importante cuya presencia apoya la validez interna del diseño. Por ejemplo, con los denominados diseños de reversión o diseños de retirada del tratamiento se puede obtener un alto grado de certeza del efecto de la intervención como agente responsable de los cambios observados en la conducta (Arnau, 1995c). Si con la retirada de la variable de tratamiento en la fase crítica se produce un cambio que se comprueba a lo largo de una serie de retiradas sucesivas (modificándose la conducta del sujeto cuando se aplica el tratamiento mientras que retorna al nivel de línea base cuando se retira dicho tratamiento) entonces se puede inferir la efectividad del tratamiento como causante de cambio.

Por último, hay una serie de creencias incorrectas respecto a los diseños de $N = 1$ que crean confusión, destacando especialmente dos:

En primer lugar, no hay que confundir los diseños de $N = 1$ con los estudios de caso tradicionales, ya que en este último tipo de estudios el análisis se lleva a cabo a través de un único sujeto, pero sus objetivos son exploratorios, no existiendo suficiente control de la situación.

En segundo lugar, en ocasiones se asocian las investigaciones que utilizan un amplio número de sujetos con análisis estadísticos mientras que las que incluyen

un único sujeto (pocos sujetos) tienen que estar asociadas necesariamente a análisis visual de los resultados. Es un grave error. Los estudios de $N = 1$ pueden ser analizados con modelos estadísticos y los que incluyen un $N > 1$ también pueden ser abordados con modelos no estadísticos como las técnicas de inspección visual (Kratochwill, 1978; Martínez, 1988).

Capítulo 5. Validez de los resultados

Marcos Pascual-Soler*

Dolores Frías-Navarro**

*ESIC Business & Marketing School, Valencia

**Universidad de Valencia

Citar el capítulo como:

Pascual-Soler, M. y Frías-Navarro, D. (2020). Validez de los resultados. En D. Frías-Navarro y M. Pascual-Soler (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> Disponible en: osf.io/kngtp

El contenido nuclear de la metodología de investigación está definido por el concepto de validez de los resultados junto con los criterios para definirla y evaluarla. En consecuencia, un programa de formación en metodología se ha de desarrollar en torno a este núcleo y sus variantes, donde la validez interna y la validez de conclusión estadística son quizá las más importantes, si nos atenemos a la necesidad de rigor, de certeza y precisión, pero sin ignorar la validez externa y la de constructo como elementos esenciales para planificar y ejecutar de forma correcta el diseño de la investigación.

La validez asociada a los resultados (a la inferencia de los resultados) de la investigación determina la calidad de los hallazgos de un estudio. Dicha calidad va a poder ser jerarquizada y no es una cuestión de todo o nada, como ya anteriormente se ha comentado.

La validez está vinculada a la calidad de las pruebas o evidencia encontrada en el estudio. Es decir, la validez está relacionada con la credibilidad de los resultados de la investigación y su aplicación a la población general de interés. Por lo tanto, un resultado es válido cuando se aproxima a la realidad verdadera del fenómeno objeto de estudio.

Planificar adecuadamente un diseño de investigación es la tarea principal del investigador o investigadora para conseguir resultados válidos. Valorar la validez de los resultados de investigación publicados es la tarea principal del lector o lectora que lleva a cabo una lectura crítica. Para una revisión completa del tema de la validez de los resultados en lengua española se pueden consultar las obras de Arnau y Balluerka (1998), Ato (1991), Balluerka (1999), García-Pérez, Frías-Navarro y Pascual-Llobell (2006) y Vallejo (1991).

Los problemas de validez, en cualquiera de sus dimensiones, no son de todo o nada, porque todas las investigaciones están expuestas a amenazas o fallos. La validez es una cuestión de grados y es una propiedad de la inferencia que se realiza a través de los resultados de la muestra a la población.

La validez no es una propiedad del diseño o de la metodología de investigación, pues un mismo diseño o la aplicación de una determinada metodología pueden producir inferencias con mayor o menor grado de validez. Por ejemplo, aplicar la metodología experimental no implica necesariamente que las inferencias sean válidas, pues factores como la pérdida selectiva de muestra o la escasa potencia de la prueba estadística podrían afectar a los hallazgos.

En definitiva, la metodología aplicada en el estudio no garantiza por sí misma la validez de una inferencia (Shadish, Cook, y Campbell, 2002).

El proceso del diseño de investigación debe garantizar la validez de los resultados y, por lo tanto, implica planificar un estudio que considere las posibles variables extrañas sistemáticas que podrían contaminar o amenazar la calidad de los hallazgos y exige controlar dichas variables para evitar su efecto sobre la variable dependiente y/o la variable independiente.

El lector o lectora consumidor de información se aproxima a la lectura de un trabajo de investigación desde la credibilidad de los hallazgos, pero su lectura siempre debe ser crítica (activa), es decir, debe valorar si los resultados que ofrece

el trabajo realmente reflejan la situación del fenómeno estudiado. Y en ese punto, la experiencia profesional y el juicio clínico ayudan a valorar la evidencia o las pruebas del estudio dentro de un contexto o diseño de investigación concreto que podría limitar la calidad de los hallazgos. Sin embargo, los conocimientos sobre metodología de investigación son necesarios si el lector realiza una lectura crítica basada en todos los elementos que forman el proceso de diseño de investigación: conocimiento previo, hipótesis, planificación, metodología, resultados y discusión. Y adelantando cuestiones que se verán con más detalle próximamente, la credibilidad de los hallazgos, su importancia y su utilidad nunca se encuentra en el valor del denominado valor p de probabilidad del resultado.

Un aspecto importante relacionado con la validez de los hallazgos tiene que ver con la ética del investigador. Afortunadamente no es la norma pero, el fraude científico existe. Se considera fraude la fabricación, falsificación y el plagio de experimentos. Muy conocido dentro de la investigación biomédica es el fraude llevado a cabo por el científico surcoreano Hwang Woo Suk, conocido como “el doctor clon”, quien en 2005 afirmaba que su equipo había obtenido células madre de embriones humanos clonados. Los resultados de sus investigaciones fueron publicados en la prestigiosa revista *Science*. Sin embargo, en el año 2006 una comisión de investigación de la Universidad de Seúl descubrió que todo era falso, concluyendo que "los datos del artículo de 2005 no fueron el resultado de simples errores, sino de la fabricación intencional", señalando que "la manipulación es un acto grave que afecta de forma negativa a los fundamentos de la Ciencia". El equipo de investigación había falsificado los datos del experimento y, además, no había conseguido esas células madre. En Octubre de 2009 el científico fue condenado a tres años de vigilancia por las autoridades.

La honestidad del científico es el eslabón que inicia el proceso de la Ciencia y representa el primer elemento de calidad de los hallazgos (Bosch, 2008).

El proceso de revisión por expertos ('revisión por pares') de los trabajos de investigación enviados a publicación, el control de las propias instituciones o de las asociaciones profesionales son claves en la corrección del fraude, el plagio y la falsificación de los hallazgos.

Existen cuatro componentes de la validez que deben planificarse en el proceso de diseño de investigación (Cook y Campbell, 1979; Shadish, Cook y Campbell, 2002):

- la validez interna
- la validez de constructo
- la validez de conclusión estadística
- la validez externa

Cuando se ejecutan investigaciones, el objetivo básico es detectar relaciones causales (si se aplica la metodología experimental) entre las variables explicativas de las hipótesis o detectar la covariación y su magnitud entre las variables (*validez interna*), así como ser capaz de captar el efecto experimental o la posible covariación de las variables si dicho efecto realmente existe en la población (*validez de conclusión estadística*) (ver Figura 15).

Además, el experimento también debe ser capaz de inferir o medir los fenómenos no observables o constructos, operacionalizando adecuadamente las variables que son objeto de estudio (*validez de constructo*) y debe proporcionar resultados que puedan ser generalizados (*validez externa*) más allá de los sujetos experimentales investigados, más allá de la propia situación del experimento y más allá del momento temporal estudiado, validez poblacional, validez ecológica y validez histórica respectivamente.

La selección aleatoria de la muestra y la replicación de los experimentos (introduciendo distintas muestras de sujetos, estímulos, tareas, mediciones...) juegan un papel destacado en el ámbito de la validez externa para confirmar la consistencia del fenómeno estudiado e incluso para idear nuevas hipótesis alternativas producto de los nuevos datos de replicación.

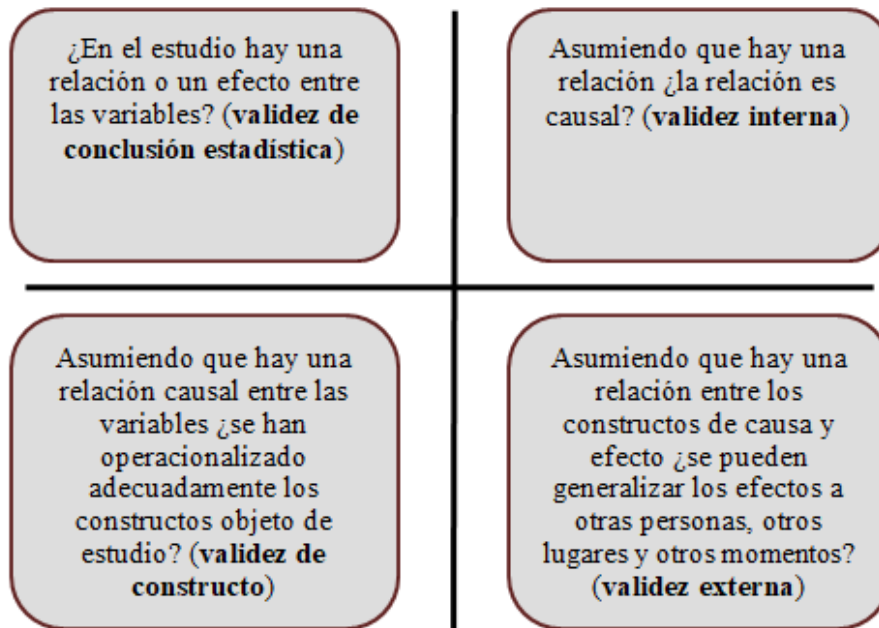


Figura 15. Validez de los resultados

Validez interna

La *validez interna* tiene que ver con la probabilidad de obtener una conclusión correcta acerca de la función de la variable independiente sobre la variable dependiente. Es decir, está relacionada con las amenazas que afectan a la inferencia causal (si se trata de una metodología experimental) o a la relación detectada entre las variables. Su propósito es asegurar que los cambios en la variable dependiente están provocados por el efecto de la o las variables independientes, si se trata de un estudio con una metodología experimental, y no por variables extrañas que intervienen en esa relación. En el resto de metodologías, la validez interna está relacionada con el efecto de sesgo que podrían producir terceras variables o variables extrañas sobre la relación no causal detectada entre las variables.

La validez interna supone valorar en qué grado los resultados obtenidos pueden ser atribuidos a la manipulación de la variable independiente (o a la asociación entre las variables) y no al efecto de variables de confundido o contaminadoras (variables perturbadoras o extrañas no controladas). Por lo tanto, la validez interna está relacionada con el control de todas las influencias sistemáticas o contaminadoras (varianza sistemática *secundaria*) que afectan a las

unidades experimentales, maximizando los efectos de la variable independiente que sí son objeto de comprobación en el estudio (varianza sistemática *primaria*) y minimizando el error aleatorio (*varianza no sistemática*).

En resumen, la validez interna está relacionada con la posibilidad de inferir conclusiones precisas y exactas sobre la relación entre las variables independientes y dependientes detectadas en la investigación.

Según la terminología de Cook y Campbell (1979), la validez interna está relacionada con la “*validez causal*”, o lo que es lo mismo, que la investigación demuestre, sin lugar a dudas, que no hay otras interpretaciones alternativas plausibles de los fenómenos objeto de estudio. Es decir, los cambios observados en la variable dependiente se deben exclusivamente a la condición definida por el investigador en la variable independiente. El problema de la validez interna es el problema de la “tercera variable” entendida como posible explicación del fenómeno: *¿existe una tercera variable (Z) responsable de la supuestamente relación evidente entre A e Y?*

Las principales amenazas a la validez interna se detallan en la Tabla 2 (Cook y Campbell, 1979).

Tabla 2. Amenazas a la validez interna

| <i>Amenaza a la validez interna</i> | <i>¿Hay una relación entre las variables independientes (A) y dependientes (Y)?</i> |
|---------------------------------------|--|
| 1. Historia | Sucesos externos al tratamiento que ocurren durante el estudio y que pueden afectar a la variable dependiente |
| 2. Maduración | Cambios biológicos y psicológicos de los sujetos que afectarán a sus respuestas |
| 3. Administración de pruebas | Los efectos del pre-test pueden alterar las respuestas en el post-test independientemente del tratamiento |
| 4. Instrumentación | Cambios en la instrumentación, o en los observadores (dificultades de calibración) |
| 5. Regresión estadística | Las puntuaciones extremas tienden acercarse a la media en el post-test a pesar del efecto del tratamiento |
| 6. Selección de la muestra | Diferencias sistemáticas en los sujetos anteriores al tratamiento |
| 7. Mortalidad experimental (atrición) | Perdida selectiva de sujetos a lo largo del estudio |
| 8. Interacciones con selección | Algunas características de los sujetos de la muestra producen errores en el efecto del tratamiento en el post-test; efectos diferenciales en los factores de selección de la muestra |

| | |
|--|--|
| 9. Ambigüedades sobre la direccionalidad de la influencia causal | Falta de claridad en la dirección de la causalidad |
| 10. Difusión/imitación de tratamientos | Los miembros de los grupos de tratamiento comparten las condiciones de tratamiento con cada uno de los demás grupos de comparación o intentan copiar el tratamiento |
| 11. Igualación compensatoria de tratamientos | Determinar que todos los sujetos tanto del grupo <i>experimental</i> como del grupo de <i>control</i> reciban un tratamiento que les proporcionen efectos beneficiosos |
| 12. Frustración de los sujetos | Los miembros de los grupos que no reciben tratamiento (grupos de control) se perciben como inferiores |

El control de las amenazas a la validez interna supone planificar los elementos que podrían afectar a las relaciones postuladas en la hipótesis científica. La técnica de control de la aleatorización es el principal elemento del diseño de investigación que si es correctamente ejecutada elimina las amenazas a la validez interna (tanto de variables observables como inobservables) y con ello las explicaciones alternativas ante los resultados obtenidos en el estudio. Por supuesto, hay otras técnicas de control dirigidas al control de las variables extrañas como la eliminación / constancia de la variable extraña, su factorización en el propio diseño de la ecuación estructural del estudio o su inclusión en la ecuación como una variable covariada, tal y como ya se ha comentado anteriormente. En la tabla 3 se detalla cómo afrontar algunas de las amenazas a la validez interna con el diseño de investigación.

Tabla 3. Control de las amenazas a la validez interna

| Amenaza a la validez interna | Control de la amenaza |
|---|--|
| Historia | Selección aleatoria, asignación aleatoria |
| Maduración | Apareamiento de los sujetos, aleatorización |
| Administración de pruebas | Grupo de control |
| Instrumentación | Fiabilidad y consistencia interna de los instrumentos |
| Regresión estadística | Omitir puntuaciones extremas, aleatorización |
| Selección de la muestra | Selección aleatoria, asignación aleatoria |
| Mortalidad experimental | Aparear a los sujetos y eliminar en ambos grupos |
| Ambigüedades sobre la direccionalidad de la influencia causal | Comprobar de forma repetida como actúa la presencia del tratamiento y su ausencia |
| Sesgos del experimentador y/o de los participantes | Enmascaramiento o cegamiento de los sujetos o de los sujetos y el experimentador (técnica de simple ciego o técnica del doble ciego) |

Desde la lógica de la investigación experimental, dos factores pueden afectar seriamente a la validez interna: el sesgo de selección y el efecto de confundido.

El “sesgo de selección” ocurre siempre que el proceso mediante el que se configuran los distintos grupos experimentales o mediante el que se asignan los tratamientos a los sujetos no garantiza que estos sean equivalentes. El sesgo de selección debe ser evitado siempre y en todos los casos, o corregido una vez recogidos los datos y antes de pasar a analizarlos, por ejemplo realizando un ajuste estadístico de los datos mediante análisis de covarianza (ANCOVA).

El sesgo de selección se puede producir en cualquiera de las fases de la investigación. En la fase del diseño y planificación se puede introducir a través de los procesos complejos de identificación de sujetos, esquema de muestreo, instrumentos de diagnóstico utilizados, etc. En la fase de implementación del experimento se puede introducir a partir de la existencia de fenómenos como la atrición o pérdida de datos. Finalmente, en la etapa de análisis de datos también se puede filtrar, por ejemplo, al analizar datos imputados y no datos reales. Desde la perspectiva del diseño y planificación de la investigación experimental, la solución universal para hacer frente al sesgo de selección es la aleatorización: mediante el proceso de asignación aleatoria del tratamiento a los sujetos o unidades experimentales (o asignación aleatoria en el orden de administración de los tratamientos si el diseño es de medidas repetidas o de medidas dependientes) garantizamos que los grupos de tratamiento o condiciones experimentales sean equivalentes entre sí previamente a la introducción del tratamiento. Gracias a la aleatorización la equivalencia de los grupos se logra tanto para variables observadas como para variables no observadas.

El concepto de “sesgo de confundido” mantienen una cierta ambigüedad y es por ello que algunos autores lo tratan al hablar de la validez de constructo y otros lo hacen al hablar de la validez interna. En principio, el efecto de confundido se comete por la manipulación inadvertida de una segunda variable (o por la covariación inadvertida de una segunda variable) o bien por la evaluación de una variable teórica relevante que no es la variable que el investigador desea estudiar. Si la manipulación de la variable independiente conlleva la covariación sistemática (inadvertida) de otra variable, estamos atentando contra la *validez interna* del experimento. Del mismo modo, si la relación entre la variable independiente y la variable dependiente depende de la presencia de terceras variables estamos también atentando contra la *validez interna*. Por el contrario, cuando se atribuye la

causalidad a un constructo teórico que no es el representado en el experimento mediante su operacionalización empírica entonces atentamos contra la denominada *validez de constructo*. La calidad de la medida (fiabilidad) realizada de los constructos o fenómenos inobservables es fundamental para garantizar la validez de constructo.

Una variable de confundido es una variable que está relacionada tanto con la variable independiente como con la variable dependiente y su presencia puede disminuir o aumentar falsamente la relación encontrada entre dichas variables de interés (Meinert, 1986). En este sentido, las hipótesis de confundido plantean que una tercera variable explica la relación entre la variable independiente y la variable dependiente

Distinguir entre los términos de mediadores y moderadores es básico para plantear el control de dichos efectos mediante el análisis estadístico y el diseño de investigación. Sin embargo, los investigadores no siempre saben diferenciar adecuadamente entre las variables mediadoras y las variables moderadoras (Ato y Vallejo, 2011; Holmbeck, 1997). La mediación no se debe confundir con una interacción o efecto de moderación donde la magnitud del efecto de A sobre Y está determinado por el valor o condición que adopte la variable moderadora. Las estrategias de análisis serán diferentes según se postule el control de una variable de confundido relacionado con el efecto de una variable mediadora o relacionado con el efecto de una variable moderadora.

Validez de conclusión estadística

Un experimento válido debe garantizar la producción del fenómeno y su generalidad, pero eso no quiere decir que, aun estando presente el efecto, el diseño de investigación tenga la capacidad de detectarlo, ya sea por el pequeño tamaño del efecto o porque el diseño no tiene la suficiente sensibilidad estadística para captar el efecto, por ejemplo por escaso tamaño de la muestra.

Se trata de la *validez de conclusión estadística* que está relacionada con el uso apropiado de las pruebas estadísticas (inferencia estadística) como instrumento para inferir si las variables del diseño covarían entre sí.

La validez de conclusión estadística tiene que ver con el componente de covariación que se establece cuando afirmamos relaciones de causa y efecto o relaciones meramente funcionales entre variables.

Se pueden plantear dos preguntas clave dentro de la validez de conclusión estadística: ¿existe covariación / correlación entre la causa y el efecto? Y ¿de qué magnitud es?

Y se pueden cometer error cuando se llevan a cabo las decisiones. Incorrectamente se puede concluir que la causa y el efecto covarían (las variables analizadas en la hipótesis de investigación covarían) cuando no es así (*error de Tipo I*) o incorrectamente concluir que no covarían cuando de hecho sí lo hacen (*error de Tipo II*).

Las dos incorrecciones anteriores están asociadas a la primera de las preguntas anteriores y hay que establecer las medidas oportunas para que no ocurran. En cuanto a la segunda pregunta, una infraestimación o una sobreponderación de la magnitud del tamaño del efecto, puede igualmente afectar a la inferencia estadística de los datos de la investigación.

Ahora ya no se trata de valorar si la relación entre las variables dependientes e independientes implicadas en la hipótesis de investigación se ha detectado sin la contaminación de posibles variables extrañas perturbadoras (validez interna). Ahora se trata de valorar si el diseño tiene las características adecuadas para detectar un efecto o una relación entre las variables, si realmente existe dicho efecto en la población.

La validez de conclusión estadística trata de garantizar que la prueba estadística empleada es la más adecuada para detectar una relación entre las variables, si dicha relación existe. Es una cuestión vinculada a si la inferencia estadística se aplica correctamente, es decir, si tiene sensibilidad la prueba estadística seleccionada por el investigador para detectar un efecto, si es la más adecuada para las hipótesis de investigación planteadas y si se aplica en las condiciones correctas relacionadas con sus supuestos estadísticos. La validez de conclusión estadística está relacionada con lo que se podría denominar el diseño estadístico de la investigación.

Controlar los denominados errores estadísticos (*error de Tipo I* y *error de Tipo II*), garantizar el cumplimiento de los supuestos de la técnica de análisis aplicada, aumentar la potencia de la prueba estadística (probabilidad de detectar un efecto si realmente existe), estimar de forma precisa el tamaño del efecto... son algunas de las cuestiones que necesariamente se deben afrontar en el apartado de la validez de conclusión estadística y que se abordarán con más detalle posteriormente al tratar el tema del procedimiento de significación de la hipótesis nula (NHST) y los diferentes diseños de investigación.

Las amenazas a la validez de conclusión estadística se detallan en la Tabla 4 (Cook y Campbell, 1979).

Tabla 4. Amenazas a la validez de conclusión estadística

| Validez de conclusión estadística | <i>¿El estudio es sensible para detectar si las variables covarían?</i> |
|--|---|
| 1. La baja potencia estadística puede provocar erróneamente un resultado estadísticamente no significativo | El error de <i>Tipo II</i> aumenta cuando el valor de alfa es bajo y la muestra es pequeña |
| 2. Violación de los supuestos de las pruebas estadísticas pueden provocar una sobreestimación o una infravaloración del tamaño del efecto y su significación estadística | Todos los supuestos de las pruebas estadísticas deben ser conocidos y comprobados cuando sea necesario |
| 3. Tasa de Error de Tipo I: repetidas pruebas estadísticas pueden aumentar la significación estadística | Se incrementa el error de Tipo I, a menos que se ajuste al número de contrastes posibles |
| 4. Fiabilidad de medición | Fiabilidad baja implica más errores que constituyen un problema serio en los estadísticos inferenciales |
| 5. Fiabilidad de la administración de los tratamientos | Los tratamientos necesitan ser administrados del mismo modo de una persona a otra, de un lugar a otro y a lo largo del tiempo |
| 6. Irrelevancias aleatorias en la situación experimental | Los efectos situacionales aleatorios pueden ser causa o interactuar con los efectos del tratamiento |
| 7. Heterogeneidad aleatoria de las unidades experimentales | Ciertas características de los sujetos pueden correlacionar con las variables dependientes |

Un aspecto muy importante vinculado a la validez de conclusión estadística es el tamaño de la muestra *¿Cuántos datos son necesarios en una investigación?*, es decir, *¿cuántos sujetos necesita el estudio para que sea sensible a captar un efecto o una relación entre las variables objeto de estudio?* Un investigador podría pensar que cuantos más sujetos mejor, pero no es una respuesta correcta. Evidentemente cuantos más datos tenga el estudio más probable será llegar a una decisión correcta dado que habrá menos error de estimación de los parámetros. Lo ideal sería tener todos los datos de la población y de este modo el investigador podría

obtener conclusiones con absoluta confianza y es más, ya no necesitaría la técnica de la inferencia estadística. Pero esto no suele ser la práctica común.

Generalmente en el mundo científico se trabaja con muestras aleatorias y se necesita de la inferencia estadística para obtener conclusiones sobre la población y por ello la estimación de la potencia estadística a priori en la fase de planificación del estudio es fundamental.

En la fase de planificación de una investigación es primordial plantearse qué tamaño muestral es necesario para garantizar de forma razonable el equilibrio entre la probabilidad de rechazar la hipótesis nula (siendo realmente verdadera, *error de Tipo I*) y el riesgo de no rechazarla (siendo realmente falsa, *error de Tipo II*). Es necesario, por lo tanto, argumentar o planificar los riesgos de error de Tipo I (seleccionar a priori el valor de alfa), de error de Tipo II (planificar un nivel aceptable de beta) y el tamaño del efecto poblacional esperado (planificar el efecto estimado). Con esas tres cantidades ya puede el investigador decidir qué tamaño muestral (N) deberá tener el estudio para garantizar la validez de conclusión estadística. Esta cuestión se tratará posteriormente cuando se analizan los diferentes índices del tamaño del efecto y el tamaño de la muestra.

En el libro de Shadish, Cook y Campbell (2002) se añade una nueva amenaza a la validez de conclusión estadística denominada “estimación imprecisa del tamaño del efecto”. Conviene tener en cuenta que algunos estadísticos sobreestiman o infraestiman de forma sistemática el tamaño de un efecto. Posteriormente al abordar la reforma estadística y la estimación del tamaño del efecto se retomará también dicha cuestión.

Validez de constructo

La *validez de constructo* alude a la relación existente entre la manipulación de la variable independiente y el constructo teórico que presumiblemente se supone que se está manipulando y también a la relación entre la variable dependiente y el constructo que se supone se está midiendo (Cook y Campbell, 1976, 1979).

Como el término constructo es sinónimo de concepto o de construcción teórica (es un invento que la comunidad científica utilizada para referirse al concepto no observable directamente, por ejemplo la ansiedad, la depresión, la atención, el

aprendizaje, el prejuicio), es evidente que la validez de constructo tiene que ver con la validez de las inferencias que se hacen acerca de los fenómenos no observables o constructos, apoyándonos en la evidencia que nos ofrecen las variables observadas que actúan como indicadores empíricos (representan una operacionalización de los constructos).

Es decir, la validez de constructo se centra en la interpretación correcta de los constructos implicados en la relación entre las variables que forman el diseño de investigación. La labor teórica de enlazar conceptos, generar hipótesis, asociar manipulaciones a variables teóricas, operacionalizar variables inobservables... es tan fundamental a la ciencia empírica como el análisis de datos.

La tarea no es fácil, dado que un mismo constructo se puede manifestar a través de múltiples variables observadas y a la vez, una misma variable puede reflejar distintos constructos. Por ello, el problema de la validez de constructo implica problemas tanto de validez convergente (relación entre las variables) como de validez discriminante (ausencia de relación entre las variables).

Las amenazas a la validez de constructo se detallan en la Tabla 5 (Cook y Campbell, 1979).

Tabla 5. Amenazas a la validez de constructo

| Validez de Constructo | <i>¿Qué variables teóricas o implícitas están siendo estudiadas?</i> |
|---|---|
| 1. Explicación preoperacional inadecuada | Escasa definición de los constructos |
| 2. Sesgo debido al empleo de operacionalizaciones únicas | Medida de una sola variable dependiente |
| 3. Sesgo debido al empleo de un método de operacionalización único. | Medida de la variable dependiente mediante un sólo método |
| 4. Adivinación de la hipótesis | Los sujetos intentan adivinar la hipótesis experimenta y actúan de la forma que creen que el investigador quiere que actúen |
| 5. Recelo de evaluación | Los sujetos manifiestan cierto recelo ante la situación de evaluación |
| 6. Expectativas del experimentador | Los experimentadores producen sesgos en el estudio a causa de sus expectativas en y durante el estudio |
| 7. Confusión entre constructos y niveles de constructo | No se implementan todos los niveles del constructo y pueden presentarse de forma débil o no existir |
| 8. Interacción de tratamientos intrasujeto | Los sujetos forman parte también de otros tratamientos (intrasujetos) |
| 9. Interacción de administración de pruebas y tratamientos | La administración de las pruebas puede facilitar o inhibir el efecto del tratamiento |
| 10. Generalidad restringida entre constructos | Punto en que un constructo puede ser generalizado de un estudio a otro |

Validez externa

El tema de la validez externa tal y como es entendida por Cook y Campbell (1976, 1979), está relacionado con la generalización de los hallazgos más allá de los sujetos del estudio, del contexto de investigación y del momento del experimento. La replicación de las relaciones halladas entre los constructos otorgará validez externa a los hallazgos.

La validez externa está relacionada con la selección aleatoria de la muestra y no se debe confundir con la asignación aleatoria del tratamiento.

La asignación aleatoria del tratamiento a los sujetos está relacionada con el control del sesgo y, por lo tanto, con la validez interna de los resultados. La selección aleatoria de las unidades experimentales (generalmente sujetos) está relacionada con la validez externa de los hallazgos.

Conviene tener en cuenta que la validez externa no es una cuestión de todo o nada. Se trata de una cuestión estrictamente empírica, pues lo que es válido para una población puede no serlo para otra. Por ello, la replicación sistemática de los hallazgos es clave para otorgar validez externa a los resultados de la investigación (Underwood y Shaughessy, 1978).

Cuando el diseño de investigación pretende comprobar teorías o buscar conductas regulares de los fenómenos entonces la replicación de los efectos elaborando nuevas investigaciones facilita el control de calidad de la validez externa. Introducir nuevas variantes del mismo fenómeno, estimar el efecto con nuevas muestras de sujetos, con nuevos estímulos y contextos permite confirmar la consistencia del fenómeno estudiado (validez interna) y al mismo tiempo permite valorar el alcance o generalización del fenómeno estudiado.

La comprobación de la consistencia de los fenómenos más allá del experimento (validez externa) también puede ponerse a prueba con la elaboración de estudios de meta-análisis.

Las amenazas a la validez externa se detallan en la Tabla 6 (Cook y Campbell, 1979).

Tabla 6. Amenazas a la validez externa

| Validez Externa | ¿Pueden generalizarse los efectos y causas de un estudio a otros sujetos, situaciones o contextos? |
|---|--|
| 1. Interacción selección-tratamiento (validez de población) | Capacidad para generalizar el tratamiento a personas que no pertenezcan al grupo estudiado |
| 2. Interacción contexto-tratamiento (validez ecológica) | Capacidad de generalización del tratamiento a situaciones más allá de la estudiada |
| 3. Interacción historia-tratamiento (validez histórica) | Capacidad para generalizar el tratamiento a otras ocasiones temporales (pasado o futuro) |

En definitiva, el diseño o plan de investigación incluye valorar los aspectos relacionados con la maximización de la *varianza sistemática primaria* (efecto del tratamiento o variable independiente), el *control de la varianza sistemática secundaria* (variables extrañas perturbadoras no controladas) y la *minimización de la variabilidad atribuida al error aleatorio* (variables extrañas aleatorias no sistemáticas), lo que supone reflexionar sobre las diferentes dimensiones de la validez del diseño de investigación que se pueden finalmente resumir en cuatro consideraciones.

En primer lugar, evaluar todos los factores que potenciarán la adecuada medición de la o las variables dependientes utilizando instrumentos de medida fiables y válidos que aseguren que se está midiendo la característica de interés de manera consistente, siendo sensible al tipo y magnitud de cambio que el investigador espera encontrar.

En segundo lugar, asegurar la manipulación correcta de la o las variables independientes que permitan que el tratamiento repercuta positivamente sobre la varianza sistemática primaria, impidiendo que esta variabilidad quede reducida por factores como la ineficaz administración de tratamientos o la dificultad (facilidad) de la tarea experimental, que de no controlarlos pueden provocar la presencia de la varianza sistemática secundaria. Dos efectos que suelen ir asociados a la variable de elaboración de la tarea experimental son los denominados efectos de techo (*ceiling effect*) y efectos de suelo (*floor effect*). El 'efecto de techo' sucede cuando la ejecución de los sujetos en la tarea es tan buena en todos los casos que resulta imposible hallar alguna diferencia entre las condiciones experimentales (por ejemplo, con tareas muy sencillas). La respuesta de los sujetos está limitada por el límite impuesto por la tarea experimental y no por las capacidades o las propias respuestas de los sujetos. En el 'efecto de suelo' la ejecución de todos los sujetos

es tan pobre que no permite discriminar entre las condiciones de tratamiento (por ejemplo, con tareas muy difíciles). En ambos tipos de efectos no se ha operacionalizado adecuadamente la manipulación de la variable independiente impidiendo observar efectos del tratamiento si los hubiese.

En tercer lugar, es objetivo del plan de investigación reflexionar sobre aquellas otras variables que no siendo teóricamente relevantes para la explicación del fenómeno objeto de interés (no se encuentran en la hipótesis del estudio), afectan de forma sistemática a la relación entre las variables explicativas del diseño al mismo tiempo que también aumentan la varianza de error del modelo. La presencia de estas variables perturbadoras debe mantenerse siempre bajo la potestad del diseño de investigación, siendo por lo tanto variables extrañas controladas.

Y en cuarto lugar, la reducción de la varianza de error no sistemático o error aleatorio (relacionada con errores de medida y diferencias individuales) constituye también un objetivo a tener en cuenta en la planificación de la investigación. La homogeneidad de las unidades de observación y la fiabilidad de los instrumentos de medida ayudarán a su disminución.

En definitiva, todas estas consideraciones tienen el propósito de obtener datos válidos (reflejo de la realidad) con los que someter a prueba la hipótesis teórica o científica (sustantiva) propuesta derivada de la necesidad de conocimiento. En este proceso, la estrategia de recogida de los datos y la naturaleza manipulada o no de la variable independiente de tratamiento marcará el tipo de metodología empleada en la investigación y la naturaleza causal o no de las relaciones encontradas, todo ello guiado siempre y primeramente por el planteamiento teórico y las hipótesis sustantivas de la investigación. La estadística está al servicio del diseño del estudio y el diseño necesita de una labor estadística realizada de forma adecuada para resolver sus hipótesis de investigación.

Capítulo 6. Resultados de la investigación (1). Diseño entre-grupos, unifactorial, univariado

Dolores Frías-Navarro*
Marcos Pascual-Soler**

*Universidad de Valencia

**ESIC Business & Marketing School, Valencia

Citar el capítulo como:

Frías-Navarro, D. y Pascual-Soler, M. y (2020). Resultados de la investigación (1). Diseño entre-grupos, unifactorial, univariado. En D. Frías-Navarro y M. Pascual-Soler (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> Disponible en: osf.io/kngtp

La herramienta de la estadística junto con la planificación del diseño de la investigación empírica que se ha realizado previamente permiten que el investigador aborde el análisis de los datos de la muestra que han sido recogidos con dicho diseño y proceder a continuación con la elaboración del informe de los resultados de la investigación.

En la pirámide del proceso de diseño de investigación se procede ahora con la fase de '*análisis de los datos*' (y se obtienen los resultados de la investigación) donde se realiza una contrastación estadística con los datos empíricos de la muestra del estudio a través del procedimiento clásico de significación de la hipótesis nula NHST.

Dicho procedimiento estadístico implica formular la *hipótesis estadística nula* (Hipótesis nula, H_0) que plantea que el efecto es cero en la población (o relación nula entre las variables del estudio; así, $H_0 \equiv \mu_1 = \mu_2$, es decir, $\mu_1 - \mu_2 = 0$).

Conviene tener muy presente, que en el proceso de contraste de hipótesis estadísticas, la hipótesis nula se asume desde el principio del análisis que es cierta y se conoce su distribución (las tablas teóricas de los estadísticos recogen sus datos según diferentes grados de libertad y nivel de significación). Gracias al conocimiento que se tiene de su distribución teórica de datos se puede comparar la probabilidad observada de un determinado resultado (empírico) asociado a un estadístico (valor p del estadístico) con la probabilidad que tendría en una distribución donde la hipótesis nula es cierta (valor teórico de las tablas o valor crítico de comparación). Y, a partir de esa decisión se procedería con la decisión dicotómica: mantener o rechazar la hipótesis nula. Por lo tanto, en el proceso de inferencia estadística solamente se trabaja con la distribución de la hipótesis nula que se asume como cierta desde el comienzo del proceso de análisis.

La *hipótesis alternativa* (Hipótesis alternativa, H_1), en cambio, plantea la existencia de un efecto diferente a cero de la o las variables independientes sobre la o las variables dependientes (sí que hay relación entre las variables estudiadas) ($H_1 \equiv \mu_1 \neq \mu_2$, es decir, $\mu_1 - \mu_2 \neq 0$) y su distribución no se valor en el proceso de contraste estadístico.

Hipótesis científica, hipótesis estadística

Tal y como se observa en la figura 16, para obtener los resultados del análisis de una hipótesis concreta es necesario realizar una serie de pasos durante el proceso del diseño de investigación.

En primer lugar, en la mayoría de las ocasiones, es necesario establecer un modelo teórico (teoría) que dará sentido al planteamiento de la hipótesis del estudio. De ahí la importancia de los dos primeros pasos de la pirámide del proceso de diseño de investigación: '*necesidad de conocimiento*' (problema a investigar) y revisión del '*conocimiento previo*' (teorías y resultados de los estudios ya realizados), ya explicados anteriormente cuando se desarrolló el proceso del diseño de una investigación.

En segundo lugar, ese conocimiento teórico permite elaborar la hipótesis científica del estudio (conocida también como hipótesis teórica o hipótesis sustantiva y formulada en términos de constructos no observables), apoyada por la revisión teórica realizada.

El planteamiento de la hipótesis científica puede ser ampliar información sobre el fenómeno objeto de estudio, introduciendo por ejemplo nuevas variables en el modelo teórico explicativo (generando conocimiento nuevo), replicar un estudio, reproducir un hallazgo previo o tratar de solucionar ciertas ambigüedades que puedan haber en ese campo de trabajo.

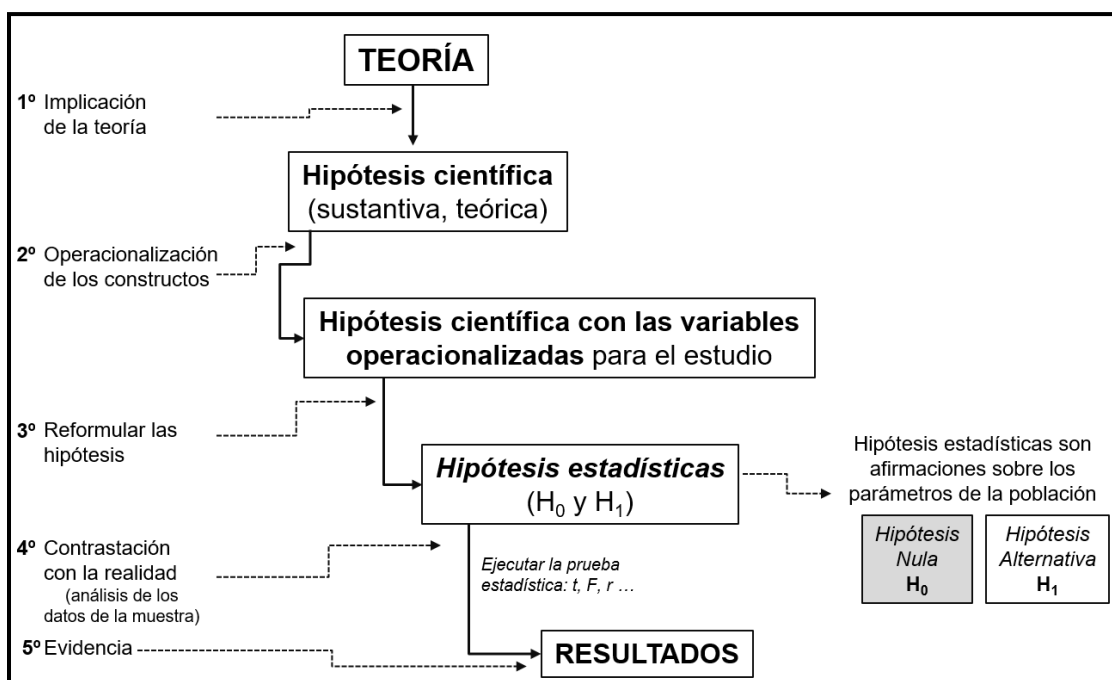


Figura 16. Hipótesis científica, hipótesis estadística y proceso de diseño de investigación

En tercer lugar, esa hipótesis teórica o científica se transforma en una hipótesis científica, pero con las variables operacionalizadas para el estudio concreto que se va a realizar. Es decir, las variables se operacionalizan para el caso concreto del estudio que se lleva a cabo (se definen las variables de forma que puedan ser medibles / observables en el estudio). Se trata de la hipótesis experimental de un estudio concreto. Así, es necesario que la hipótesis científica se transforme para ser sometida a su estudio empírico y, para ello, el investigador operacionaliza los constructos de la hipótesis sustantiva en variables que pueden ser medidas para

poder captar en la realidad cómo se comporta el fenómeno que está siendo objeto de estudio (hipótesis científica operacionalizada).

Por lo tanto, conviene tener muy presente que todas las variables del estudio (independientes, dependientes y extrañas controladas) tiene 2 dimensiones:

- como constructo teórico no observable y
- operacionadas en el diseño para poder ser medidas y recoger los datos del estudio o experimento

Por ejemplo, si el investigador desea analizar si la indefensión aprendida provoca depresión (se trata de dos constructos no observables) será necesario que en el diseño del estudio el investigador operacionalice esos constructos para poderlos medir y comprobar la relación entre ellos. Podría operacionalizar el constructo de la indefensión con una tarea experimental donde un determinado efecto se mostrará haga lo que haga el sujeto. No hay control de la situación por parte del sujeto y, por lo tanto, se producirá indefensión aprendida. Y, además, el investigador tendrá que medir la depresión a través de, por ejemplo, un autoinforme o una entrevista que mida la presencia del trastorno de forma fiable y válida o el tiempo invertido en recorrer un laberinto. En definitiva, es necesario operacionalizar o transformar los enunciados de las hipótesis científicas en enunciados que puedan ser contrastados empíricamente con la realidad. Y, ese proceso de transformación necesariamente implica que el investigador describa de forma detallada en el informe de investigación la calidad de sus decisiones, aportando información sobre la fiabilidad y consistencia interna de la medida de los constructos (variable dependiente) así como argumentar por qué la tarea experimental (variable independiente) representa al fenómeno objeto de estudio. Por lo tanto, la hipótesis teórica de la indefensión aprendida provoca depresión podría ser operacionalizada como: “las ratas sometidas a una situación de indefensión aprendida tardarán más tiempo en recorrer un laberinto”.

En cuarto lugar, se procede al análisis de los datos de la muestra y con ello a la contrastación del modelo teórico con la realidad. Se trata de ejecutar el denominado ‘contraste de hipótesis estadísticas’. En el procedimiento tradicional de contraste de hipótesis se asume, como principio, que la hipótesis nula es cierta (en la población

el efecto o la relación entre las variables es cero) y gracias a conocer su distribución teórica (distribución teórica o muestral del estadístico) se podrá llegar en el estudio a una conclusión estadística al comparar la probabilidad del resultado obtenido con la muestra (valor p del estadístico utilizado en la prueba estadística ejecutada en el estudio) con la probabilidad que tendría dicho resultado en esa distribución de H_0 . En la distribución de la hipótesis nula se asume un efecto de cero o ausencia de relación entre las variables (distribución teórica o muestral del estadístico que se encuentra en las tablas de los manuales de estadística y de diseño de investigación).

La hipótesis nula es la protagonista del contraste de hipótesis y solamente ella interviene al tomar la decisión estadística: se *mantiene* la hipótesis nula o se *rechaza* la hipótesis nula (decisión estadística dicotómica). No se debe decir 'se acepta la hipótesis nula', ya que, desde el principio del contraste estadístico, se acepta dicha hipótesis como cierta.

En ese proceso de contraste estadístico, la hipótesis alternativa tiene un papel secundario, ya que sólo entra en acción cuando se rechaza la hipótesis nula, ya que, como consecuencia, se aceptará la hipótesis alternativa de efecto diferente a cero.

Cuando se mantiene la hipótesis nula, tras finalizar la decisión dicotómica, no se debe interpretar la hipótesis alternativa ni concluir que no hay efecto o no hay relación entre las variables, pues en ese caso se trata de un resultado no concluyente y, en todo caso, se necesita seguir investigando sobre él. Es un error muy común creer que mantener la hipótesis nula significa que las medias son iguales, que las puntuaciones no correlacionan o que no hay efecto del tratamiento. Y no es así.

Sin embargo, desde el punto de vista del diseño de la investigación y la Práctica Basada en la Evidencia, el investigador o investigadora no debe centrar todos sus esfuerzos únicamente en 1) decidir si mantiene o no la hipótesis nula (significación estadística), pues es fundamental llevar a cabo una interpretación más allá del valor p , recurriendo al 2) análisis e interpretación del tamaño del efecto y su intervalo de confianza (significación del tamaño del efecto), contextualizando la magnitud de los efectos hallados dentro del área de investigación del fenómeno

objeto de estudio, 3) a la interpretación clínica o sustantiva del hallazgo, basada en el juicio del experto o juicio clínico, respecto a su utilidad dentro de los objetivos del estudio (significación clínica) y, 4) también, a la calidad del diseño de investigación que se ha ejecutado, llevando a cabo una valoración crítica del control efectuado con el diseño del estudio sobre las amenazas a la validez interna, externa, de constructo y de conclusión estadística (valorar conscientemente y de forma juiciosa la calidad de la evidencia hallada).

Por otra parte, el lector o lectora activo que lleva a cabo una lectura crítica del informe de investigación también debe poner en marcha esas reflexiones sobre la técnica y el análisis metodológico que se ha llevado a cabo en el artículo o informe y valorar todas las facetas de la significación de los hallazgos y no solo la faceta de la significación estadística. Gracias a esa valoración, el lector o lectora puede decidir qué grado de validez tienen las conclusiones de ese trabajo de investigación.

Lograr que un resultado sea estadísticamente significativo no significa de forma directa que dicho resultado es válido y fiable, pues hay más causas que podrían llevar a dicha significación (si fuese así se habría cometido un error estadístico Tipo II o un falso positivo). Ni tampoco significa que sea un resultado importante, ni tampoco útil. E incluso, a veces, el investigador o investigadora puede dar un resultado como estadísticamente significativo cuando se puede leer claramente en el texto que no es así (se trataría de un grave error de educación estadística por parte del investigador) y, de ahí, la importancia de leer críticamente o de forma activa los resultados del trabajo o informe y no dejarse llevar solamente por las conclusiones que ha redactado el autor o autora del informe.

De alguna manera, todos los lectores y lectoras activos se convierten en meta-investigadores, es decir, deben llevar a cabo una lectura rigurosa y crítica que valore cada apartado del informe, interpretando de nuevo el diseño del estudio y, si es necesario, utilizando herramientas o programas que le permitan comprobar los resultados que se aportan en el trabajo o añadir resultados que el investigador o autor del artículo no ha incorporado en su estudio.

Por ejemplo, si el autor del artículo no lo ha realizado en su informe, podría ser interesante estimar el tamaño del efecto y su intervalo de confianza y ampliar la

interpretación de la significación estadística (valor p) con la aportada por la significación del tamaño del efecto y valorar la incertidumbre de su estimación puntual gracias a la interpretación de su intervalo de confianza. También, el juicio clínico del investigador podría acompañar a los resultados del estudio y, por ejemplo, en el apartado de discusión del artículo o informe se podría debatir los hallazgos desde un punto de vista sustantivo y/o aplicado y no solamente desde la significación estadística y del tamaño del efecto obtenido.

Por lo tanto, en ocasiones, puede ser necesario que el mismo lector activo lleve a cabo una estimación de otros indicadores que no han sido incluidos en el artículo o informe, pero sería interesante disponer de dicha información para comprender y valorar mejor los hallazgos y su alcance.

Por todo ello, es muy importante que el estudio primario presente un informe transparente y con toda la información fundamental de los estadísticos básicos. Por ejemplo, es esencial informar siempre, al menos, del número total de observaciones que se ha empleado en el estudio (N), número de observaciones utilizadas en cada condición experimental (n), medias y desviaciones típicas de todas las puntuaciones mencionadas en el estudio y de las pruebas de contraste estadístico efectuadas junto con sus características como los grados de libertad y el valor de p exacto del estadístico calculado. Toda esa información es imprescindible cuando se redacta el informe.

Además, disponer de un informe transparente y completo ayuda a realizar los procesos de revisión de los artículos ya publicados por otros investigadores desde un punto de vista crítico (se trata de actividades de meta-investigación: investigar sobre la investigación publicada) y, con ello se potencia la calidad de la Ciencia, se promueve que la Ciencia pueda autocorregirse y se detectan casos de fraude y de prácticas de investigación cuestionables. La acumulación del conocimiento científico necesita del filtro del control de los miembros de la propia comunidad científica.

Los denominados estudios de meta-investigación han demostrado que en la literatura publicada hay más errores en los diseños de investigación de lo que sería deseable y muchos problemas de comprensión y educación estadística entre los investigadores o autores de los artículos e informes. Y, si a estas circunstancias se

une el hecho de que hay problemas de redacción y de transparencia en los artículos e informe, entonces la tarea de la lectura crítica se hace mucho más complicada y necesaria.

En la Figura 17 se detallan unos resultados del artículo de Van Dessel y De Houwer (2019) titulado “Hypnotic suggestions can induce rapid change in implicit attitudes”. En este momento sería interesante que leyera esos resultados y tratará de ir más allá de las palabras que redactan los investigadores. Trate de sacar sus propias conclusiones leyendo de forma activa los resultados que aportan los investigadores. O, quizás, aun no puede llevar a cabo esa lectura crítica porque no dispone del conocimiento metodológico necesario. El análisis de los datos de esta figura 17 se retomará posteriormente cuando se haya presentado el contraste de hipótesis para observar cómo informar y redactar los resultados de los estadísticos.

Experiment 2

Implicit evaluation. Scores for the pre- and postmanipulation AMP were computed by subtracting the percentage of “pleasant” responses on trials with the negative-induction person from the percentage of “pleasant” responses on trials with the positive-induction person. An ANOVA on AMP scores revealed a main effect of time, $F(1, 68) = 35.07, p < .001, \eta^2 = .34, BF_{10} > 1,000$, and a marginally significant interaction effect of time and hypnosis, $F(1, 68) = 3.52, p = .065, \eta^2 = .05, BF_{10} = 1.09$. Planned contrasts did not reveal a significant difference between the hypnosis condition ($M = 0.25, SD = 0.28$) and relaxation condition ($M = 0.25, SD = 0.38$) at Time 1, $t(70) = -0.01$, one-tailed $p = .50$, 95% CI for the mean difference = $[-\infty, 0.13]$, $d = 0.00$, 95% CI for $d = [-0.23, 0.23]$, $BF_{01} = 4.50$. In contrast, and most crucially, at Time 2, AMP scores were lower in the hypnosis condition ($M = -0.24, SD = 0.43$) than in the relaxation condition ($M = -0.01, SD = 0.40$), $t(70) = -2.42$, one-tailed $p = .009$, 95% CI for the mean difference = $[-\infty, -0.07]$, $d = 0.41$, 95% CI for $d = [0.17, 0.65]$, $BF_{10} = 4.57$. Interestingly, AMP scores were reduced from Time 1 to Time 2 in both the hypnosis and relaxation groups, $t_s < -3.01, p_s < .005, BF_{10}s > 17.08$.

Figura 17. Ejercicio de lectura crítica de los resultados del informe

En estos momentos este ejercicio es solamente de reflexión y de autoevaluación personal. Pero es importante tener en cuenta que trabajar siguiendo el planteamiento de la Práctica Basada en la Evidencia exige tener los conocimientos de metodología de investigación suficientes para leer de forma crítica o activa los artículos e informes que se publican dentro de su área de trabajo.

Además, desde el punto de vista ético, es necesario seguir formándose en las nuevas técnicas de análisis que se van incorporando en los informes o en los nuevos objetivos de los estudios como puede ser leer de forma crítica las revisiones sistemáticas, los trabajos de meta-análisis y de meta-análisis en red, el denominado Factor Bayes, el tamaño del efecto y su intervalo de confianza y comprender adecuadamente los fundamentos de los procedimientos clásicos de análisis como las pruebas de contraste estadístico tipo t de Student, razón F , coeficiente de correlación o análisis de regresión.

El Factor Bayes (BF)

Antes de continuar con el tema, solamente un apunte breve sobre el denominado Factor Bayes (BF) y su interpretación, ya que su uso en las Ciencias del Comportamiento aun no se ha generalizado, pero poco a poco va apareciendo más en los artículos e informes.

El Factor Bayes informa sobre la probabilidad de la hipótesis nula respecto a la probabilidad de la hipótesis alternativa, dados unos datos, $p(H|\text{datos})$. O viceversa: la probabilidad de la hipótesis alternativa respecto a la probabilidad de la hipótesis nula, dados unos datos. Es decir, se trata de una interpretación totalmente diferente al valor p de la perspectiva frecuencial que tradicionalmente se aplica en las pruebas de inferencia estadística basadas en la comprobación de la hipótesis nula (NHST).

El valor p de la inferencia frecuencial (procedimiento clásico de la hipótesis nula) informa de la probabilidad de unos datos, o datos más extremos, dada la hipótesis nula como cierta ($p(\text{datos}|H_0)$), pero no dice nada de la probabilidad que tiene una hipótesis estadística de ser cierta (ni tampoco de ser falsa).

El valor del Factor Bayes, en cambio, se puede expresar como apoyo a la hipótesis nula respecto a la alternativa y se representa

1. BF01 expresses the likelihood of H0 relative to H1 given the data.
2. BF01 expresses the probability of the data given H0, relative to H1.

como BF_{01} o, en cambio, su valor puede indicar el apoyo de la hipótesis alternativa sobre la hipótesis nula y se representa como BF_{10} .

El Factor Bayes proporciona un continuo de evidencia para la probabilidad de H_1 sobre la de H_0 , o viceversa.

Un valor de $BF = 1$ significa que los datos apoyan de la misma manera a la hipótesis nula que a la hipótesis alternativa (no hay evidencia que apoye a ninguna hipótesis). Es decir, los datos predicen de la misma manera a ambas hipótesis estadísticas y la evidencia no favorece a un modelo sobre el otro.

A medida que se incrementa el valor del BF, aumenta la probabilidad de una hipótesis estadística sobre la otra hipótesis. Una de las interpretaciones que más se utilizan del Factor Bayes es la propuesta por Jeffereys (1961) que señala que si FB_{10} (apoyo de la hipótesis alternativa sobre la hipótesis nula) es:

- > 100 entonces hay extrema evidencia para H_1
- Entre 30-100 evidencia muy fuerte para H_1
- Entre 10-30 fuerte evidencia para H_1
- Entre 3-10 hay evidencia moderada para H_1
- Entre 1-3 es evidencia anecdótica para H_1
- Si el $BF = 1$ entonces no hay evidencia para ninguna de las dos hipótesis estadísticas.

Cuando se calcula FB_{01} (evidencia a favor de la hipótesis nula sobre la hipótesis alternativa) se utilizan los mismos valores anteriormente descritos para llevar a cabo la interpretación de FB. Sin embargo, conviene ser precavido y no caer en la tentación de que esos valores sean la única guía para valorar la calidad de los hallazgos y, de nuevo, se conviertan en un oráculo de la verdad tal y como se ha criticado duramente al valor de punto de corte de .05 en la perspectiva frecuencial que fue establecido por Fisher. Siempre es necesario recurrir a valorar todo el proceso del diseño de investigación y jerarquizar la evidencia o pruebas que aporta el estudio, actuando siguiendo la perspectiva de la Práctica Basada en la Evidencia. El análisis de todos los elementos que forman parte del proceso del diseño de investigación es la herramienta que facilita la valoración activa de la validez de los resultados, y no solamente los resultados estadísticos.

En definitiva, Jeffereys (1961) sitúa el valor de 3 como punto de referencia para considerar que hay evidencia moderada relevante para ser interpretada. Y no es un valor arbitrario, ya que, en general, está vinculado a un resultado estadísticamente

significativo ($p < .05$). Sin embargo, hay que anotar que hay debate sobre el valor de referencia para indicar que hay evidencia sustantiva y otros investigadores / revistas optan por un valor de 6 como ocurre en la revista *Cortex* o de 10 tal y como se señala en la revista *Nature Human Behavior for Registered Reports*. Actualmente, y desde el área de la perspectiva de la inferencia clásica frecuencial, también hay debate sobre si seguir manteniendo el valor de .05 como punto de corte para decidir si hay o no significación estadística, pues hay opiniones de expertos que consideran que debería ser menor, como por ejemplo .005.

El programa JASP, gratuito, permite calcular el Factor Bayes de una forma sencilla. Para realizar los cálculos del Factor Bayes se parte de que las dos hipótesis tienen a priori la misma probabilidad (50%) y el BF expresa la probabilidad de los datos bajo los dos modelos, H_0 y H_1 . Por ejemplo, un $BF_{10} = 4$ significa que los datos son 4 veces más probables que ocurran bajo la hipótesis alternativa que bajo el modelo de la hipótesis nula H_0 , por lo tanto, $BF_{01} = 1/4$). O, por ejemplo, $BF_{10} = 32$ significa que los datos son 32 veces más probables que ocurran bajo el modelo de la hipótesis alternativa (probabilidad de 32 : 1 a favor de H_1) y, ese dato se corresponde de forma directa con un valor de $BF_{01} = 1/32$ (probabilidad de 1 : 32 a favor de H_0).

Resultados nulos

Continuando con la exposición del contraste de hipótesis desde la perspectiva clásica de inferencia estadística basada en el modelo frecuencial (modelo de comprobación de la significación de la hipótesis nula, NHST), también es importante tener en cuenta que los resultados donde se mantiene la hipótesis nula (conocidos como 'resultados nulos' y también como 'resultados negativos') deben ser publicados, llevando a cabo una interpretación de su valor de significación estadística, pero también del tamaño del efecto estimado y de su intervalo de confianza (que permite valorar la incertidumbre en la estimación puntual) junto con la apreciación de todos los elementos metodológicos del diseño que podrían favorecer ese resultado nulo como la baja potencia estadística, el escaso tamaño de la muestra, la estimación a priori excesiva del tamaño del efecto, explicaciones al tamaño del efecto detectado

Un problema muy importante que tiene la Ciencia actual es que aproximadamente el 80%-90% de los resultados que se publican son resultados que han sido estadísticamente significativos, 'olvidándose' en el cajón del archivador aquellos estudios donde se mantiene la hipótesis nula (se trata del problema del denominado 'sesgo de publicación'). Como consecuencia de esa conducta del investigador, el panorama de resultados que ofrece el mundo científico es demasiado bueno y no se ajusta a la realidad de los fenómenos. Por ello, hay que luchar desde el mundo académico y profesional para dar valor a los resultados nulos (se mantiene la hipótesis nula) que también tienen su papel en la explicación de la realidad detectada e implican reflexionar sobre por qué se producen o qué elementos del diseño han estado amenazados y su valoración requiere, en gran medida, la estimación del tamaño del efecto y su intervalo de confianza. Y desde el modelo de análisis NHST no se pueden interpretar nunca como evidencia de ausencia o evidencia de relación, pues ausencia de evidencia no es evidencia de ausencia.

En definitiva, los resultados nulos que se producen al mantener la hipótesis nula tras ejecutar la prueba estadística ($p > .05$) no significa que no hay efecto en la población o que las variables no están relacionadas; solamente significa que, con los resultados obtenidos, no se puede llegar a una conclusión sobre la hipótesis teórica propuesta. Se trata de un resultado no concluyente desde el punto de vista de la significación estadística, pero que acompañado del tamaño del efecto y su intervalo de confianza y con la valoración de la calidad del diseño aplicado se pueden plantear ideas para mejorar la próxima investigación y seguir de este modo indagando sobre la naturaleza y características del fenómeno objeto de estudio.

Es muy importante dar a conocer los resultados nulos y publicarlos porque su ausencia enturbia la visión que se tiene de un fenómeno e impide seguir estudiando líneas de investigación apoyadas en los esfuerzos que realizaron otros investigadores, pero cuyo trabajo se desconoce, ya que nunca lo publicaron.

Por otra parte, si lo que el investigador desea comprobar en su hipótesis teórica es que dos medias no difieren de forma estadísticamente significativa o que no hay una relación entre dos variables (se plantea una hipótesis de igualdad) entonces su

técnica de contraste de hipótesis son las denominadas “pruebas de equivalencia” y no el procedimiento tradicional de contraste estadístico.

El procedimiento tradicional de contraste de hipótesis estadística NHST está dirigido a analizar hipótesis teóricas de superioridad donde los grupos difieren, ya que las puntuaciones de un grupo serán superiores a las del otro grupo aunque, como ya se ha comentado, todo el contraste se base en la hipótesis de nulidad de efectos o ausencia de relación entre las variables (H_0). Pero esa cuestión es una de las paradojas que envuelven al contraste estadística.

Una vez ejecutados los análisis estadísticos, se obtienen unos resultados (evidencia) cuyo grado de validez depende de forma directa de la calidad del proceso de diseño de investigación que se haya planificado y realizado en el estudio. En el diseño debe primar la *maximización de la varianza primaria* que se atribuye a la relación entre las variables que forman la hipótesis (variables independiente y dependiente) y *minimizar la varianza no sistemática o error aleatorio* presente siempre en el modelo. Por supuesto, la calidad o validez de los resultados también depende del control efectivo de la *varianza sistemática secundaria* (variable extraña) cuyo efecto no controlado en el diseño invalidaría las conclusiones del estudio.

Ecuación estructural del modelo

Para comprobar un modelo teórico es necesario formularlo como un modelo matemático (ecuación estructural) y contrastarlo con datos empíricos, realizando un proceso de modelización estadística. El objetivo de dicha modelización es separar la variación que se atribuye al componente sistemático del modelo, efecto = α , (vinculado a los efectos de la variable de tratamiento; varianza sistemática primaria) de la variación relacionada con el componente aleatorio o término de error del modelo, error = ε (varianza no sistemática o aleatoria). Incorporando en el modelo, si fuese necesario, las fuente de varianza sistemática secundaria que se han controlado: por ejemplo, un factor de bloques o una variable covariada.

En este punto es muy importante definir adecuadamente los componentes del modelo lineal (ecuación estructural) que se pretende ajustar a los datos obtenidos con el estudio. De ahí la importancia de revisar y construir de forma adecuada el

modelo teórico, pues el modelo matemático que se formule estará determinado por el planteamiento teórico.

El Modelo Lineal General (MLG, representado en la ecuación estructural del diseño) es un modelo estadístico que describe una combinación lineal de los efectos aditivos que forman la puntuación obtenida en la denominada variable dependiente $Y = \text{constante, efecto y error}$.

En las Ciencias del Comportamiento la estimación de la variable dependiente no está exenta de error, utilizándose modelos estadísticos o probabilísticos a diferencia de los modelos determinísticos donde se plantean relaciones exactas entre las variables. El término ε representa el término de varianza aleatoria, es decir, recoge el efecto de otras variables no incluidas directamente en el modelo y cuyo efecto individual no debe ser ni destacado ni sistemático (debe ser error aleatorio).

Diseño de grupos independientes unifactorial univariado

La ecuación estructural o modelo matemático que plantea el Modelo Lineal General (MLG) que describe a un diseño de grupos independientes (conocido como *diseño entre-grupos* o *diseño entre-sujetos*), con una sola variable independiente (diseño unifactorial) y una sola variable dependiente (diseño univariado) se basa en descomponer la puntuación de la variable independiente Y en tres partes:

1) *Constante* o media general (M).

2) *Efecto* o efectos del tratamiento o varianza explicada por el efecto de estar en un grupo u otro (se representa como $A, B, C...$ y se trata de la denominada 'varianza entre-grupos').

3) *Error* o varianza explicada por el efecto que se produce dentro de cada grupo (efecto intra-celdilla) que debe estar vinculado con las posibles fuentes de variabilidad no sistemática como el error aleatorio vinculado al muestreo o las propias diferencias individuales (E , se trata de la denominada 'varianza-intragrupo' o varianza intra-celdilla). Es decir, los sujetos que reciben una misma condición de la variable independiente (todos se encuentran por ejemplo en el grupo a_1) tendrán

el mismo efecto vinculado a dicha condición (α_1), pero cada sujeto tendrá un término de error de estimación diferente (de ahí el término de variabilidad o varianza intra-grupo), pues a pesar de recibir el mismo tratamiento a_1 no todos los sujetos obtienen la misma puntuación en la variable dependiente Y. En este punto, hay que tener muy en cuenta que la fuente de varianza o de variabilidad del término de error también podría incluir los efectos de otras variables con efecto sistemático no controladas en el diseño (variables contaminadoras sin control). La validez de los resultados del estudio (validez de conclusión estadística) depende de que ese término de error solamente incluya variabilidad no sistemática o aleatoria y que el efecto de las variables de varianza sistemática secundaria hayan sido controladas de forma efectiva en el diseño de investigación. Si el término de error contiene variables extrañas entonces estará sesgado, ya que su valor será mayor y, por lo tanto, afectará a la validez de conclusión estadística, pues se pierde potencia estadística al efectuar el contraste de hipótesis.

La representación de la ecuación estructural en términos poblacionales de un *diseño entre-grupos unifactorial univariado*, en términos del modelo completo o saturado o de la hipótesis alternativa, es la siguiente:

$$Y = \mu + \alpha + \varepsilon$$

Donde Y representa las puntuaciones en la variables dependiente medida (los datos del estudio), μ es la constante o media total de los datos, α es el efecto de la variable independiente o factor (que tendrá al menos dos condiciones: α_1 y α_2 , pero podría tener tres o más condiciones) y ε es el término de error que debe ser aleatorio y que se encuentra en los datos vinculado el error de muestreo y/o diferencias individuales.

En términos de matrices, el modelo o ecuación estructural anterior se representa como:

$$Y = M + A + E$$

Donde, tal y como ya se ha descrito, Y es la variable dependiente, M es la constante o media general, A es el efecto del tratamiento y E es el error aleatorio. Este tipo de representación de matrices es la que se utiliza en este material para el

desarrollo de las ecuaciones estructurales de cada modelo de diseño de investigación.

Por ejemplo, en un *diseño entre-grupos unifactorial univariado* los sujetos tendrán una sola medición o puntuación, ya que formarán parte de una condición u otra (entre-grupos), hay una variable independiente (A) que tiene dos condiciones (a_1 y a_2) y una variable dependiente (Y) y la ecuación estructural plantea que las puntuaciones en la variable dependiente Y de cada uno de los sujetos que forman la muestra puede descomponerse en:

1. la media general del grupo o constante,
2. más el efecto de estar en una determinada condición del factor o variable independiente (un efecto para los sujetos del grupo o condición a_1 y otro efecto para los sujetos del grupo a_2 , donde la suma de los efectos siempre será cero, ya que la estimación del Efecto es igual a: $A = M_a - M$),
3. más el error aleatorio no sistemático, propio de la puntuación de cada sujeto, ya que se vincula a las diferencias individuales o errores de muestreo; (Error $E = Y - M - A$; la suma de los valores de los errores intraceldilla o dentro de cada grupo también sumará siempre cero).

Como ya se ha comentado, es un requisito fundamental que el diseño controle con alguna técnica de control las variables extrañas o contaminadoras porque su efecto estará en el término de error, aumentando su varianza, y, por lo tanto afectando de forma directa al valor del estadístico de la razón F del Análisis de la varianza (ANOVA) que disminuirá al estar contaminado el valor de la varianza que se considera aleatoria por factores de varianza sistemática secundaria (problemas de potencia estadística; amenaza a la validez de conclusión estadística).

Una vez que se ha definido la ecuación estructural o modelo matemático que representa a la hipótesis del estudio y recogidos los datos, se procede con la fase de ejecución del *análisis de los datos*.

Una de las técnicas estadísticas que más se utiliza para analizar el efecto de la variación atribuida al efecto del tratamiento respecto a la que procede del error aleatorio es el denominado Análisis de la Varianza (ANalysis Of VAriance, ANOVA) que utiliza la prueba estadística de la Razón F para llevar a cabo el contraste

estadístico. Este tipo de modelos o pruebas de ANOVA serán analizadas en el presente libro.

Análisis de la varianza (ANOVA)

La técnica del Análisis de la Varianza (ANOVA) permite comparar dos o más medias entre sí para determinar mediante el proceso de contraste de hipótesis estadísticas si la diferencia (o las diferencias) entre las medias de la variable dependiente en los grupos de la variable independiente es estadísticamente significativa. Para ello, la ecuación estructural plantea un modelo matemático o ecuación estructural donde la variable dependiente Y es la variable cuantitativa continua y la o las variables independientes (factores del modelo: A, B, C ...) son las variables cualitativas o factores que estarán formados por diferentes condiciones que representarán diferentes efectos (efectos de las condiciones a_1 y a_2 : α_1 y α_2 ; efectos de las condiciones b_1 y b_2 : β_1 y β_2 ; y efectos de las condiciones c_1 y c_2 : χ_1 y χ_2 ..., por ejemplo).

La denominada Razón F es el estadístico o prueba de contraste que se utiliza en el análisis de la varianza.

En los diseño de grupos independientes o diseño entre-grupos, la Razón F compara la 'varianza entre los grupos' (efecto de la variable independiente en cada condición o grupo) con la 'varianza intra-grupo' (error aleatorio o varianza aleatoria que se produce dentro del grupo). Si se concluye que la diferencia entre la varianza entre-grupos y la varianza intra-grupos es estadísticamente significativa (el valor p del resultado obtenido con la prueba estadística de la razón F es menor a .05 dentro del modelo de la distribución de la hipótesis nula) entonces al menos una de las medias es diferente de forma estadísticamente significativa de al menos otra media de la variable dependiente. De ahí el nombre de análisis de la varianza, ya que compara la varianza entre-grupos con la varianza intra-grupo para detectar diferencias entre las medias de los diferentes grupos o condiciones que forman el diseño de la investigación.

$$F = \frac{\text{Varianza del efecto}}{\text{Varianza del error}}$$

En resumen, la técnica de ANOVA permite conocer el vínculo de los cambios en la variable dependiente según las condiciones o niveles de las variables independientes. Es decir, permite analizar si las puntuaciones en la variable dependiente difieren o no de forma estadísticamente significativa cuando se comparan los resultados de cada grupo o condición de la variable independiente. Cuando el diseño tiene más de dos grupos, puede ser que haya una sola diferencia estadísticamente significativa entre dos grupos o puede que hayan más diferencias entre las medias de los grupos, pero el ANOVA no especifica el número de grupos que difieren ni qué grupos difieren, ya que se trata de una prueba 'omnibus' y solo informa de que hay alguna diferencia estadísticamente significativa entre los grupos. Por lo tanto, si el diseño tiene más de dos grupos ($A > 2$, tiene más de una diferencia de medias), será necesario continuar el análisis de los datos con otras técnicas que permitan descubrir entre qué grupos se encuentran las diferencias estadísticamente significativas como las 'pruebas de contraste de hipótesis específicas', también conocidas como pruebas post hoc o pruebas a posteriori (por ejemplo, Tukey, Bonferroni...).

Modelos de la hipótesis nula y la hipótesis alternativa

En la figura 18 se representa la formulación de la ecuación estructural de un *diseño entre-grupos unifactorial univariado* bajo el planteamiento de la hipótesis nula y de la hipótesis alternativa. Se observa que en el modelo de la hipótesis nula o modelo restringido, la descomposición de la puntuación en la variable dependiente solamente incluye la constante o media general (M , que será la puntuación predicha o puntuación pronostica por el modelo para cada sujeto del estudio) más el término de error ($Y = M + E$).

En cambio, cuando se trata del modelo completo o modelo de la hipótesis alternativa, la ecuación estructural sí tiene el efecto del tratamiento o el efecto de estar en un grupo u otro (A) junto con la constante o media general y el error ($Y = M + A + E$). En este modelo de la hipótesis alternativa la puntuación predicha por el modelo para cada sujeto será la constante más el efecto que haya producido estar en ese grupo (puntuación predicha $\hat{Y} = M + A$). Por lo tanto, los sujetos que se encuentren en el mismo efecto tendrán el mismo efecto de A y la suma de los

efectos de todos los grupos siempre será igual a 0 ya que la estimación de los efectos son desviaciones respecto a un valor medio (estimación del efecto de $A = M_a - M$). La suma del cuadrado de los efectos estimados es igual a la Suma de Cuadrado del efecto de A ($\Sigma(A)^2$).

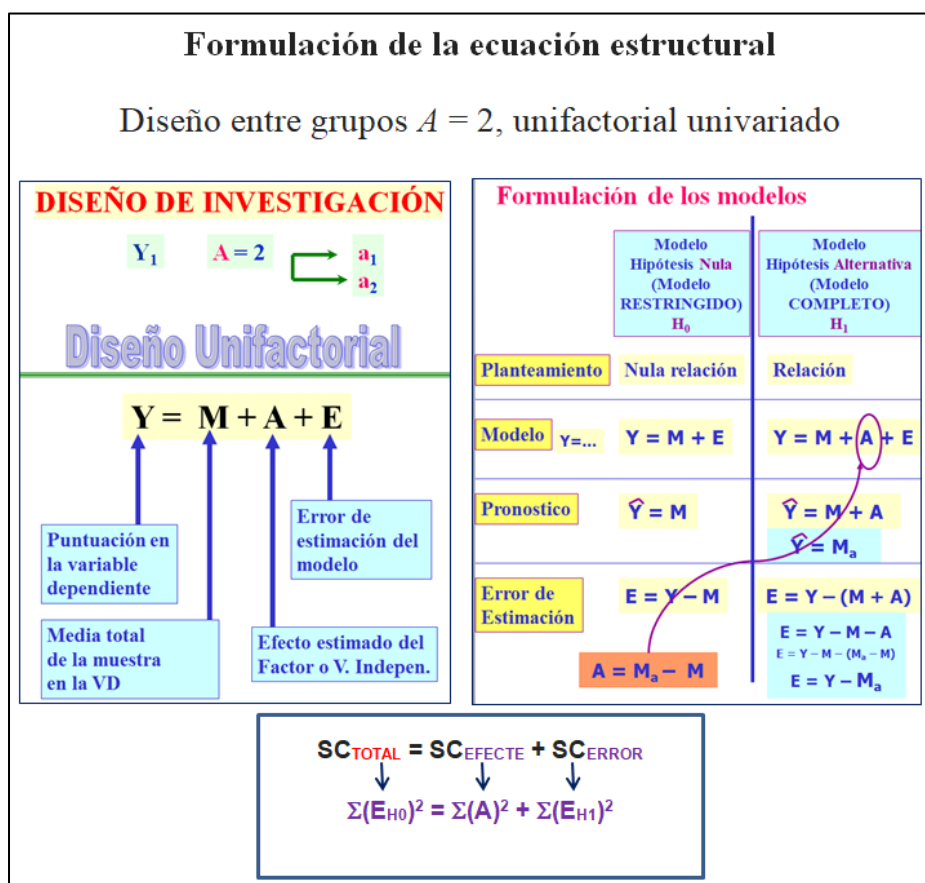


Figura 18. Formulación de los modelos: hipótesis nula / hipótesis alternativa

Cada modelo estadístico plantea un pronóstico o una puntuación pronosticada para cada uno de los sujetos que forman la muestra. En el caso del modelo de la hipótesis nula que señala que no hay relación entre las variables, la puntuación que pronostica para todos los sujetos es la media general o constante. Es decir, dado que no hay efecto de la variable independiente, el pronóstico de las puntuaciones solamente se basa en el valor de la media general de todas las puntuaciones. En este caso, el error de estimación es la diferencia entre la puntuación obtenida en el estudio (dato en la variable Y) menos la puntuación pronostica que es la media general ($E = Y - M$). La suma del cuadrado de los errores estimados es igual a la Suma de Cuadrado del error del modelo completo o de la hipótesis alternativa ($\Sigma(E_{H1})^2$).

En el modelo de la hipótesis alternativa (plantea que hay relación entre las variables) se introduce la fuente de la varianza del efecto de la variable y , por lo tanto, la puntuación que pronostica para cada sujeto será la constante o media general más el efecto de estar en un determinado grupo del diseño, tal y como ya se ha comentado. En este caso, el error de estimación es la diferencia entre la puntuación obtenida en el estudio Y menos la puntuación pronostica por el modelo en su ecuación estructural: $E = Y - M - A$.

El componente del error de estimación del modelo o ecuación estructural siempre es la puntuación obtenida en el estudio menos la puntuación pronostica por dicho modelo. Así, una vez formulada la ecuación estructural resulta sencillo calcular dicho error como $E = Y - \hat{Y}$ por el modelo. Recordar que \hat{Y} es la puntuación en la variable dependiente Y pronosticada.

La variabilidad total o varianza total observada en las puntuaciones Y será igual a la varianza explicada por el efecto de la variable independiente (SC, Suma de Cuadrados del efecto) más la varianza explicada por el término de error (SC, Suma de Cuadrados del error). La Suma de Cuadrados Total es el igual a la suma de los cuadrados de los valores del error de estimación del modelo nulo ($\sum(E_{H_0})^2$). Posteriormente, se detallará la explicación del concepto de sumas de cuadrados.

Para desarrollar la ecuación estructural se puede utilizar la plantilla de la figura 19 donde se especifica cada uno de los pasos necesarios para ejecutar el ANOVA y se explica cómo estimar cada uno de los elementos que forman la descomposición de la ecuación estructural hasta llegar a la tabla del análisis de la varianza o tabla de ANOVA (figura 19) (ver también Anexo 2).

| Fuentes de Varianza (FV) | Sumas de Cuadrados (SC) | Grados de Libertad (gl) | Medias Cuadráticas (MC) | Razón F (F) | Valor de Probabilidad (p) | Tamaño del Efecto (η^2) |
|--------------------------|-------------------------|-------------------------|-------------------------|-------------|---------------------------|--------------------------------|
|--------------------------|-------------------------|-------------------------|-------------------------|-------------|---------------------------|--------------------------------|

alfa = _____; $F_t(.05, gl_A, gl_E) =$ _____; $F_e(gl_A, gl_E) =$ _____; $p =$ _____

| | | | | | | |
|--|--|--|--|--|--|--|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

Figura 20. Tabla del ANOVA y decisión estadística

Estimación de los parámetros: término del efecto de A

Una vez identificado el modelo estadístico que representa a la hipótesis de investigación (identificados en una ecuación estructural), a continuación es necesario estimar los parámetros que componen la ecuación estructural.

Los parámetros se pueden definir como ‘efectos principales’, como ‘efectos simples’, ‘efectos de interacción’, ‘error’ o incluso como ‘contrastes concretos’.

La estimación del parámetro del efecto principal del factor A (variable independiente) se define como la media del grupo o condición donde se encuentra el sujeto menos la media poblacional, midiéndose por lo tanto en la escala de puntuaciones de diferencia:

$$\alpha = \mu_{ai} - \mu$$

Es decir,

$$A = M_a - M$$

Estimación de los parámetros: término de error

La estimación numérica del parámetro del error mediante el criterio de los ‘mínimos cuadrados’ supone obtener un estimador que minimice el componente aleatorio del modelo, ε . Se trata de obtener la minimización de la suma cuadrática

de los errores, de manera que la suma de los errores de estimación del modelo debe ser la mínima posible: Es decir, la predicción de los valores de la variable dependiente que realiza el modelo estadístico formulado debe incluir el menor error de estimación posible.

El error o residual $\hat{\varepsilon}_i$ estimado del modelo o ecuación estructural refleja la diferencia entre el valor de la puntuación obtenida en la variable dependiente Y y el valor predicho por el modelo para la puntuación estimada en dicha variable \hat{Y} .

$$\hat{\varepsilon} = Y - \hat{Y}$$

Y la puntuación predicha de la variable dependiente \hat{Y} se estima como:

$$\hat{Y}_i = M + \text{ TODOS LOS EFECTOS FORMULADOS EN EL MODELO}$$

Así, en el caso de un diseño con un solo factor (factor A), la puntuación predicha de la variable dependiente \hat{Y} se estima como:

$$\hat{Y} = M + A$$

Por lo tanto, la estimación del error se efectúa como:

$$\hat{\varepsilon} = Y - (M - \text{ TODOS LOS EFECTOS FORMULADOS EN EL MODELO})$$

En el caso de un diseño con un solo factor (factor A), el error $\hat{\varepsilon}$ se estima como:

$$\hat{\varepsilon} = Y - (M - A)$$

Luego,

$$\hat{\varepsilon} = Y - M - A$$

Y, en este tipo de diseño entre-grupos unifactorial, se puede comprobar que si $A = M_a - M$, entonces, $\hat{\varepsilon}_i = Y - M - (M_a - M)$, y por lo tanto:

$$\hat{\varepsilon} = Y - M_a$$

Razón F

El valor del estadístico F del ANOVA es la razón entre la variación de los datos observada entre los distintos niveles o condiciones de la variable independiente (efecto de A) respecto a la varianza que se atribuye al error (efecto de E).

$$F = \frac{\text{Varianza del efecto}}{\text{Varianza del error}}$$

Suma de Cuadrados

Para valorar la parte de variación atribuible al efecto del tratamiento A respecto a la variabilidad aleatoria de los datos (E) hay que considerar todas las observaciones (puntuaciones en la variable dependiente Y) que se han registrado en la investigación. El problema estriba en que como el modelo se expresa en puntuaciones de diferencia, si se suman todos los valores (de todos los sujetos) de algún componente de la ecuación, el resultado siempre será cero (pues son desviaciones respecto a la media). Esta medida permite que el usuario compruebe si está realizando bien los cálculos, pues si la suma de los efectos estimados no da 0 o la suma de los errores de los sujetos de la muestra no da 0 entonces es que se ha realizado la estimación de dichos efectos de manera incorrecta.

Por lo tanto, si se suman los efectos del tratamiento ($\hat{\alpha}$) de las N observaciones de la muestra de un estudio el resultado de la suma siempre será cero ($\sum \hat{\alpha} = 0$). Del mismo modo ocurre con los valores del componente de error que suman cero al sumar todos los errores. Y, también, se puede comprobar que suman cero al sumar los valores de error de los sujetos que están dentro de la misma condición de A, por ejemplo a_1 . Y, también suman cero las puntuaciones de diferencia (y) que representan cada puntuación directa en Y menos la constante o media general. Debido a ello, el procedimiento para sumar las puntuaciones de diferencia de cada fuente de variación se realiza elevando al cuadrado estas puntuaciones y posteriormente sumándolas ($\sum (\hat{\alpha})^2$). Con ello se evita que los datos sumen cero y se proporciona un valor de suma de cuadrados.

A la puntuación que resulta de sumar el cuadrado de las puntuaciones de diferencia de una fuente de variación se le denomina “Suma de Cuadrados” de

dicha fuente de variación (Suma de Cuadrados del efecto o variable independiente A, Suma de Cuadrado del error E, Suma de Cuadrados Total T).

Así, la suma de cuadrados es una transformación cuadrática de las puntuaciones de diferencia que mantiene la proporción respecto de las diferencias originales, eliminando el signo de la puntuación y con ello se evita que siempre sume cero. De tal forma que si se extrae la raíz cuadrada del cuadrado, se obtiene la puntuación de diferencia en valores absolutos. El cuadrado permanece invariante ante el signo y aumenta en proporción cuadrática a la distancia que separa la media de la observación.

Grados de Libertad y Medias Cuadráticas

Como cada una de estas sumas de cuadrados (de los efectos y del error) se estima con un número distinto de observaciones independientes, se debe corregir estas diferencias dividiendo cada suma de cuadrados por sus correspondientes “grados de libertad”.

Los grados de libertad indican el número de observaciones cuyos valores son libres de variar, o en otras palabras, el número de observaciones independientes de una fuente de variabilidad menos el número de parámetros estimados al computar dicha variación. Por ejemplo, los grados de libertad de la suma de cuadrados total son $N - 1$.

Al resultado de dividir la suma de cuadrados de una fuente de variación por sus grados de libertad se le denomina “Media Cuadrática” de esa fuente de variación.

La prueba F es la razón entre las medias cuadráticas del efecto A y el error E y siempre que se escriba un valor de F hay que informar de los grados de libertad de las dos fuentes de varianza que se han utilizado para obtener el valor de ese estadístico. Se trata de la razón $F_{empírica}$ del estudio:

$$F(\text{gl efecto}, \text{gl error}) = \frac{\frac{SC A}{gl A}}{\frac{SC ERROR}{gl ERROR}} = \frac{MC A}{MC ERROR}$$

SC es la Suma de Cuadrado y MC es la media cuadrática; gl es los grados de libertad.

Los grados de libertad del efecto del tratamiento A (varianza 'entre-grupos') se obtiene como número de condiciones que tenga la variable independiente menos 1 ($gl_A = a - 1$).

Los grados de libertad del error (varianza 'intra-celdilla' o del error) se obtiene como número total de observaciones menos el número de condiciones que tenga el diseño ($gl_{\text{error}} = N - a$).

El estadístico F mide el número de veces que es mayor la suma de cuadrados del efecto de un determinado tratamiento (efecto dividido por sus grados de libertad), que la fuente de error del modelo, (dividido también por sus correspondientes grados de libertad). La prueba de la hipótesis del estadístico F compara varianzas; de ahí que el nombre que reciba esta prueba, tal y como se ha comentado, sea precisamente el de análisis de la varianza.

Decisión estadística

Si los resultados obtenidos con la investigación confirman o no las hipótesis científicas o teóricas se comprueba calculando la probabilidad de obtener el tamaño del efecto detectado con la muestra del estudio en una distribución muestral de una hipótesis estadística que plantea que el efecto es cero (modelo de la hipótesis nula).

Como ya se ha comentado, conviene tener muy presente que con el contraste de hipótesis estadísticas no se comprueba la probabilidad de que la hipótesis nula sea cierta o falsa porque se parte como supuesto o condición que la hipótesis nula es cierta y se conoce su distribución muestral donde se asocia cada valor con un valor de probabilidad bajo dicho supuesto.

Dado que la hipótesis nula generalmente plantea que el efecto es cero en la población entonces todos los valores de su distribución tendrán la misma probabilidad, siendo la distribución muestral uniforme (ver figura 21).

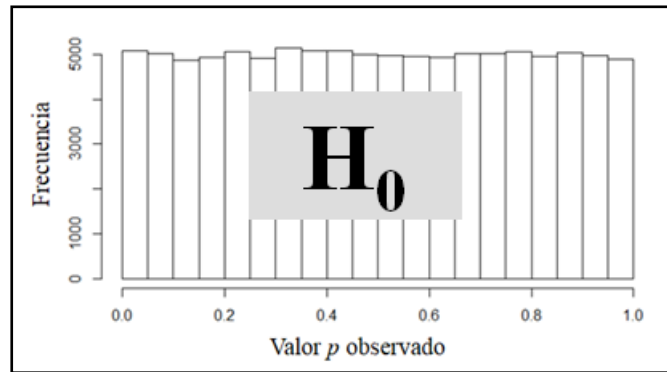


Figura 21. Distribución de los valores p de las frecuencias en el modelo de la hipótesis nula

Teniendo en cuenta esa cuestión, el contraste estadístico calcula el valor p del estadístico obtenido con los datos de la muestra (se conoce como $F_{empírica}$), asumiendo que la hipótesis nula es cierta. Y, para tomar la decisión estadística (mantener la hipótesis nula o rechazar la hipótesis nula) compara dicho valor p del estadístico con el que tendría en la distribución de la hipótesis nula ($F_{teórica}$), dados los grados de libertad 'entre' o del efecto y los grados de libertad 'intra' o del error. Así, para llevar a cabo esa comparación se compara el valor de la $F_{empírica}$ con el valor de la $F_{teórica}$ que se puede encontrar en la tabla de la distribución muestral o teórica de la hipótesis que se encuentra en los manuales de estadística y diseño, existiendo diferentes tablas en función del valor del alfa (.05, .01, .025...) donde se consulta según los grados de libertad asociados a la prueba estadística (ver figura 22).

TABLAS Razón F

Tabla III (continuación). $F (\alpha = 0.050, gl_{entre} = columnas, gl_{error} = filas)$

| gl | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 24 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 161.448 | 199.500 | 215.707 | 224.583 | 230.162 | 233.986 | 236.768 | 238.883 | 240.543 | 241.882 | 243.906 | 249.052 |
| 2 | 18.513 | 19.000 | 19.164 | 19.247 | 19.296 | 19.330 | 19.353 | 19.371 | 19.385 | 19.396 | 19.413 | 19.454 |
| 3 | 10.128 | 9.552 | 9.277 | 9.117 | 9.013 | 8.941 | 8.887 | 8.845 | 8.812 | 8.786 | 8.745 | 8.639 |
| 4 | 7.709 | 6.944 | 6.591 | 6.388 | 6.256 | 6.163 | 6.094 | 6.041 | 5.999 | 5.964 | 5.912 | 5.774 |
| 5 | 6.608 | 5.786 | 5.409 | 5.192 | 5.050 | 4.950 | 4.876 | 4.818 | 4.772 | 4.735 | 4.678 | 4.527 |
| 6 | 5.987 | 5.143 | 4.757 | 4.534 | 4.387 | 4.284 | 4.207 | 4.147 | 4.099 | 4.060 | 4.000 | 3.841 |

Figura 22. Tabla de la razón F para un alfa = .05, grados de libertad del efecto en las columnas y los grados de libertad del error en las filas

Y cuando el valor de la $F_{empírica} \geq F_{teórica}$ entonces se rechaza la hipótesis nula, ya que el valor de p vinculado al estadístico de la $F_{empírica}$ será menor que el valor de alfa que se haya utilizado al consultar las tablas de los manuales (generalmente alfa es .05 y, por lo tanto, $p \leq .05$ se considera que el efecto o la relación detectada entre las variables es estadísticamente significativa).

Cuando se consultan las tablas teóricas de la distribución muestral del estadístico (construidas bajo el supuesto de la hipótesis nula) se necesita conocer:

1) el valor de alfa que se ha establecido a priori (antes de recoger los datos) como riesgo asumido de error de Tipo I por parte del investigador o investigadora en la fase de planificación del estudio y

2) los grados de libertad (gl) que se utilizan en el estadístico o prueba estadística seleccionada por el investigador para dar una solución empírica a su hipótesis científica.

En el caso de la razón F el estadístico compara dos varianzas que tienen diferentes grados de libertad y, por lo tanto, se necesitará conocer los grados de libertad de la fuente de varianza del efecto (fuente de varianza entre-grupos) y los grados de libertad del error (fuente de varianza intra-grupo) para obtener el valor de la $F_{teórica}$. En el caso de la prueba t de Student (solamente se puede utilizar si el diseño tiene solamente 2 grupos, $A = 2$), únicamente se necesitará conocer los grados del término de error, pues los del tratamiento o efecto siempre serán 1 porque el diseño solamente tiene dos condiciones (los grados de libertad del tratamiento se obtienen como número de condiciones menos 1; $a - 1$, luego $2 - 1 = 1$). Los grados de libertad del error se obtienen como $N - a$. Y los grados de libertad totales son $N - 1$.

Errores estadísticos

El contraste estadístico siempre se realiza asumiendo un determinado riesgo de equivocación al rechazar una hipótesis nula que realmente es cierta (se trata del “nivel de significación o valor de alfa”, α , o “error de Tipo I”: probabilidad de rechazar la hipótesis nula siendo realmente cierta. El nivel de significación está

establecido por consenso en el 5% ($\alpha = .05$) si no se indica otro valor en el informe del estudio.

A la probabilidad complementaria del error de Tipo I se le denomina “nivel de confianza” ($1 - \alpha$) y determina cuál será la probabilidad de que el investigador acierte en su decisión cuando la hipótesis nula sea realmente cierta, es decir, cuando se mantiene la hipótesis nula y realmente no existe relación entre las variables independientes y las dependientes especificadas en la hipótesis de investigación.

Cuando se rechaza la hipótesis nula, se podría estar cometiendo un error de Tipo II (error beta, β). Se trata de la probabilidad de mantener la hipótesis nula siendo realmente falsa. A su probabilidad complementaria se le denomina potencia estadística ($1 - \beta$).

La potencia estadística es una decisión correcta y es la la probabilidad de rechazar la hipótesis nula siendo realmente falsa. Es decir, la probabilidad de detectar un efecto si realmente existe en la población. Generalmente se establece que el error de Tipo II debe ser de .2 como mucho, es decir, se considera como aceptable más error de tipo II ya que sería cuatro veces mayor que el error de Tipo I.

En resumen, el contraste estadístico se inicia partiendo del supuesto de que no existe relación entre la variables independiente y la dependiente (modelo de la hipótesis nula). El modelo de la hipótesis nula atribuye las diferencias (ya sean las diferencias grandes o sean pequeñas) que pudiesen aparecer entre las puntuaciones de los diferentes grupos al efecto de extracción aleatoria de las muestras (muestreo aleatorio) y a las diferencias individuales no sistemáticas. Es decir, cualquier diferencia, grande o pequeña, detectada entre las condiciones es aleatoria (fruto del azar) porque en la población el efecto es cero.

Cuando se ejecuta la prueba del contraste de hipótesis estadísticas (por ejemplo con la razón F , con la t de Student, con r del coeficiente de correlación...), si se detecta que algún efecto de los postulados en el modelo estadístico no es un efecto nulo o de efecto cero entonces la hipótesis nula se puede rechazar, ya que al menos una de las condiciones experimentales (o alguna posible combinación entre varias de ellas o efecto de interacción) no procede de la misma población.

Cuando se rechaza la hipótesis nula se puede aceptar la hipótesis alternativa que está vinculada con el efecto postulado en la hipótesis científica o teórica si se tienen garantías de la validez de la evidencia aportada por los resultados. Ante el rechazo de la hipótesis nula hay un abanico de posibles hipótesis alternativas, pero si el diseño se planifica de forma adecuada, controlando las amenazas a las valideces, entonces cabe suponer que el efecto obtenido está provocado por las causas propuestas en el modelo teórico del estudio. Si el diseño de investigación es deficiente, evidentemente a pesar de rechazar la hipótesis nula poco se podrá decir de la causa de los efectos hallados.

En resumen, una vez se determina el tamaño del efecto experimental con los datos del estudio, se calcula su probabilidad bajo el supuesto de la hipótesis nula. Se puede concluir que el resultado es estadísticamente significativo (valor p de probabilidad obtenido con la prueba estadística \leq valor de alfa fijado a priori por el investigador) y rechazar la hipótesis nula. Por el contrario, puede ocurrir que se mantenga la hipótesis nula si el valor de probabilidad asociado con el tamaño del efecto detectado supera el punto de corte de α preestablecido ($p > \alpha$). Para llevar a cabo dicha decisión estadística se puede emplear el estadístico razón F del Análisis de la Varianza cuyo proceso de ha detallado anteriormente.

Además, conviene tener muy claro que solamente el tipo de metodología que se haya empleado en el estudio podrá fundamentar que las hipótesis planteen hipótesis de causalidad donde la variable independiente causa un efecto sobre la variable dependiente. Este tipo de hipótesis de causalidad solamente pueden plantearse cuando se trata de un diseño de investigación que se ha planificado con una metodología experimental (manipulación de la variable independiente y asignación aleatoria del tratamiento). En otros contextos de investigación donde se emplea la metodología cuasi-experimental (manipulación de la variable independiente sin asignación aleatoria del tratamiento) será necesario poner en marcha un control exhaustivo de posibles diferencias previas entre los sujetos que podrían contaminar los resultados por terceras variables que no estaban presentes en la hipótesis del estudio y se podría hablar de cuasi-causal. Y, por supuesto, si el estudio se lleva a cabo con una metodología no experimental (ni manipulación de la variable independiente, ni asignación aleatoria del tratamiento) nunca se deben plantear las hipótesis como causales ni interpretar los hallazgos como que la

variable independiente causa un efecto sobre la variable medida, pues solo se interpretará en términos de covariación entre las variables y magnitud de la relación.

En definitiva, el tipo de metodología empleada en el estudio determinará el alcance de las inferencias de los resultados del análisis de la varianza que será de causalidad o de asociación entre las variables explicativas (independiente y dependiente) en función de la metodología de investigación empleada en el estudio.

Supuestos estadísticos del modelo

La prueba estadística aplicada es completamente válida en la medida que se cumplan los supuestos fundamentales de las pruebas paramétricas: que la variable dependiente se distribuya normalmente, que las varianzas de los grupos sean homogéneas (se puede aplicar la prueba de Levene para comprobar este supuesto donde se requiere que el resultado del contraste estadístico sea $p > .05$) y que las distintas observaciones sean independientes y extraídas aleatoriamente, no existiendo entre las condiciones experimentales ninguna otra diferencia sistemática que la atribuida al efecto de la variable independiente.

Los supuestos de las pruebas paramétricas señalan que debe existir:

Normalidad: los errores (ε_{ij}) sigue una distribución normal. Esto es equivalente a que Y_{ij} sigue una distribución normal.

Linealidad: $E(\varepsilon_i) = 0$. Esto es equivalente a que $E(Y_{ij}) = \mu + \alpha_i + \beta_j$.

Homocedasticidad de las varianzas: $\text{Var}(\varepsilon_{ij}) = \sigma^2$. Esto es equivalente a que $\text{Var}(Y_{ij}) = \sigma^2$.

Independencia: ε_{ij} son independientes entre sí. Esto es equivalente a que Y_{ij} son independientes entre sí.

En general, los estudios de simulación han comprobado que la razón F es robusta al incumplimiento de dichos supuestos si la muestra es lo suficientemente grande, siendo un problema si la muestra es pequeña.

Desarrollo de la ecuación estructural. Ejemplo 1

SUPUESTO DE INVESTIGACIÓN. Supongamos que un investigador desea comprobar si la indefensión aprendida produce déficits depresivos. Diseña una situación experimental donde los sujetos son ocho ratas ($N=8$) que deben completar un laberinto de cuyo suelo reciben descargas eléctricas de baja intensidad ininterrumpidamente. La mitad se asigna aleatoriamente a la condición de shock escapable (a1) y la otra mitad a la condición de shock inescapable (a2). La tarea experimental consiste en recorrer el laberinto, midiéndose el tiempo (en segundos, Y) que emplean las ratas en su recorrido. La hipótesis experimental mantiene que las ratas del grupo de shock inescapable fracasarán en su aprendizaje, recorriendo el laberinto con lentitud y sin precisión, e incluso en muchas ocasiones no llegaran a completarlo, manifestando un comportamiento depresivo. Sin embargo, las ratas que se encuentran en la condición experimental de shock escapable aprenderán que con su ejecución escapan de la descarga y aumentarán rápidamente la velocidad de carrera con objeto de eliminar la situación aversiva y llegar al habitáculo que les privará de las descargas, a diferencia de la condición de shock inescapable donde perdurará la descarga aunque lleguen a dicho habitáculo. Tras una serie de diez ensayos previos que facilitaron el aprendizaje de la situación, los resultados del experimento fueron los que se detallan en la Tabla 7.



Tabla 7. Matriz de resultados

| A → Shock | Y → Tiempo |
|-----------------|----------------|
| a1 Escapable | 23, 11, 12, 26 |
| a2 No escapable | 39, 38, 23, 28 |

Las características de la investigación planteada en el ejemplo anterior permiten concluir que se trata de un diseño de investigación con una metodología experimental con dos condiciones de la variable independiente A donde su manipulación y la asignación aleatoria del tratamiento a las unidades de

observación (sujetos) indica que se trata de un experimento verdadero o realizado con una metodología experimental.

De acuerdo con la hipótesis que se quiere comprobar se formula la ecuación estructural, conocida también como modelo estructural paramétrico, donde se realiza una transformación lineal de los datos separando los dos componentes del modelo (efecto y error) cuya relación lineal determina que los efectos sean aditivos:

- 1) el efecto de las variables independientes (componente fijo o de efecto fijo para cada sujeto de la condición) y
- 2) el término de error (componente aleatorio vinculado a la puntuación de cada sujeto).

Una vez especificado el modelo del diseño se pondrán a prueba una serie de hipótesis de nulidad relacionadas con los diferentes efectos atribuidos a las variables independientes de la ecuación estructural; en el ejemplo únicamente es una variable independiente A con dos condiciones o niveles de tratamiento.

La estimación del efecto del shock (variable independiente A) vendrá determinada por la diferencia en tiempo que se observe entre las dos situaciones o condiciones manipuladas (a_1 : shock escapable; a_2 : shock inescapable). Si la única diferencia constante que ha existido entre las dos condiciones, a lo largo de todo el experimento, es la posibilidad o no de escape del shock, cabría esperar que las medias de los dos grupos fueran distintas. Claro, en la medida que la hipótesis teórica de partida sea cierta.

El modelo se suele expresar en variaciones en torno a la media. La manera más simple e intuitiva de escribir el modelo estructural consiste en expresar cada puntuación individual de Y como la suma de la media del grupo del sujeto más el término de error específico de cada sujeto. Se trata de la ecuación estructural:

$$Y = \mu + \alpha + \varepsilon$$

Es decir:

$$Y = M + A + E$$

Pudiéndose estimar los parámetros del efecto como la media del grupo respecto a la media poblacional, midiéndose por lo tanto en la escala de puntuaciones de diferencia:

$$\alpha = \mu_a - \mu$$

Es decir,

$$A = \mu_a - M$$

Donde se deduce que:

$$\mu_a = \mu + \alpha$$

Es decir,

$$\mu_a = M + A$$

Como estos son los valores de la población de ratas y en el ejemplo se trabaja con una muestra, lo que se obtiene es una estimación de estos parámetros poblacionales desconocidos a partir de las medias muestrales, donde los parámetros $\hat{\alpha}$ representan las desviaciones de las medias marginales respecto a la media general:

$$\hat{\alpha} = \bar{y}_a - \bar{y}$$

Dado que la suma de estas desviaciones es igual a cero, como ya se comentó, entonces el último parámetro de las condiciones (α_2) se puede obtener como diferencia. Aplicando esta fórmula a los datos del supuesto se obtiene que $\alpha_1 = -7$, $(18 - 25)$, por lo tanto como $\sum_{j=1}^k \alpha_j = 0$ entonces necesariamente $\alpha_2 = 7$. De este modo, el número de parámetros que se necesita estimar en el modelo son tantos como el número de condiciones experimentales $-a-$ menos uno (Riba, 1990).

El cambio producido en la escala cuando se trabaja con los efectos de las condiciones, respecto a sus medias, no varía la información, sino que únicamente expresa los datos tomando como punto de referencia la media. Por tanto, si conocemos la media de los datos y la puntuación de diferencia, podemos conocer el tiempo que se ha invertido en el recorrido del laberinto. Hay dos consecuencias

fundamentales que se derivan de este primer cambio de escala para el análisis de los resultados:

1) Se calcula el valor de un parámetro, la mediat total \bar{Y} (M), que indica el punto central de todas las puntuaciones.

2) La inclusión de un parámetro implica la pérdida de un grado de libertad en el conjunto de los datos. Esto quiere decir que los N valores obtenidos en el experimento se conocen a partir de $N - 1$ puntuaciones de diferencia y del valor del parámetro de media total, M (Box, Hunter, y Hunter, 1978, página 40).

Una vez se conocen los parámetros del modelo y se formula la ecuación estructural, se pueden expresar los datos sobre su base y de este modo comprobar la predicción del modelo. En el ejemplo, en cualquiera de las dos situaciones de shock, la puntuación que se pronostica para la velocidad del recorrido resultará de sumar a la constante o media general M, el efecto correspondiente al nivel del factor donde se encuentra el sujeto:

$$\hat{Y} = M + A$$

Y el error (ε) cometido en la observación i será la diferencia que resulte de restar a la puntuación real (Y) el pronóstico que se obtenga con la ecuación del modelo (\hat{Y}_j):

$$\varepsilon = Y - \hat{Y}$$

Por ejemplo, en la observación cuarta - Y_4 - se registraron 26 segundos en el recorrido del laberinto. Como esta unidad experimental se encontraba en la condición de shock escapable (a_1) el pronóstico o puntuación pronosticada para ese sujeto (y para todos los sujetos que se encuentren en su misma condición o grupo: a_1), será:

$$\hat{Y}_{a_1} = \bar{Y} + A_1 = 25 + (-7) = 18 \text{ segundos}$$

Como realmente esa rata (Y_4) ha recorrido el laberinto en 26 segundos, el error de estimación de esa concreta rata número 4 (cada sujeto o rata tiene su propio error) ha sido:

$$\varepsilon_4 = Y_4 - \hat{Y}_{a1} = 26 - 18 = 8 \text{ segundos}$$

La razón última del modelo es expresar la hipótesis en términos mensurables para medir el efecto que ha tenido en la variable dependiente los cambios operados en la variable independiente.

Como los datos se expresan en puntuaciones de diferencia se obtiene que la Suma de Cuadrado Total ($\Sigma(Y - M)^2$) es igual a la Suma de Cuadrado del Efecto ($\Sigma(\hat{Y} - M)^2$) más la Suma de Cuadrados del error ($\Sigma(Y - \hat{Y})^2$):

$$Y - M = (\hat{Y} - M) + (Y - \hat{Y})$$

Luego, la diferencia de 2 segundos por debajo de la media en el tiempo invertido en el recorrido del laberinto de la rata 4 se corresponde con un descenso de 7 segundos por la posibilidad de escapar del shock al final del recorrido y de 5 por el efecto de otras variables distintas de la manipulada. De este modo, el modelo descompone la variación respecto de la media total en dos partes: una atribuida al tratamiento o no del shock, y otra a las restantes variables que inciden en la velocidad del recorrido. Siendo estas últimas la fuente de variación de error por proceder de variables que no están contempladas en el modelo teórico.

Si se suman los efectos del tratamiento ($\hat{\alpha}$) de las N observaciones de la muestra el resultado de la suma será cero. Por ello, es necesario calcular las denominadas Sumas de Cuadrados de cada fuente de variación para obtener un valor de varianza o variabilidad en cada fuente de varianza.

El procedimiento para sumar las puntuaciones de diferencia de cada fuente de variación se realiza elevando al cuadrado estas puntuaciones y sumándolas. A la puntuación que resulta de sumar el cuadrado de las puntuaciones de diferencia de una fuente de variación se denomina "Suma de Cuadrados" de dicha fuente de variación.

Por tanto, la Suma de Cuadrados correspondiente a la manipulación del factor A o explicada por la acción del tratamiento (Suma de Cuadrados del efecto o del tratamiento) en el ejemplo que nos ocupa es igual a 392.

La suma de las puntuaciones de diferencia de los datos respecto a la media general al cuadrado ($\sum Y - M$) se denomina Suma de Cuadrados Total. En el ejemplo la Suma de Cuadrados total es igual a 748.

Y, la suma de cuadrados correspondiente a las variables no incluidas en el modelo se conoce como Suma de Cuadrados del Error. En el ejemplo la Suma de Cuadrados residual o del error es igual a 356.

Las sumas de cuadrados se descomponen por lo tanto en las puntuaciones de diferencia de cada observación. De tal forma que:

$$SCTOTAL = SCA + SCERROR$$

Luego la descomposición de la varianza total de las puntuaciones del ejemplo propuesto es la siguiente:

$$SCTOTAL = SC_{EFECTO DEL SHOCK} + SC_{ERROR} = 392 + 356 = 748$$

Durante los primeros momentos del aprendizaje del desarrollo de la ecuación estructural de un modelo de diseño se pueden seguir las indicaciones de una plantilla de aprendizaje tal y como se ha señalado anteriormente. Para este ejercicio se puede consultar en el Anexo 2 la plantilla que se ha elaborado para el aprendizaje de un diseño entre-grupos unifactorial univariado. A continuación se detalla en la figura 23 el inicio del ejercicio utilizando dicha plantilla.

| α (condición: $a_1, a_2, a_3 \dots$) | S (sujeto: 1, 2, 3, 4, 5...) | Y (puntuaciones en la variable dependiente) | M (Media general o constante) | y ($Y - M$) E_{H0} | A ($M_a - M$) | \hat{Y} ($M + \text{efectes}$) Puntuación pronosticada H1 | E ($Y - \hat{Y}$) $E_{H1} = Y - M_a$ |
|--|---|---|--|-------------------------------------|---------------------------|--|---|
| a1 escapable | 1 | 23 | | | | | |
| a1 escapable | 2 | 11 | | | | | |
| a1 escapable | 3 | 12 | | | | | |
| a1 escapable | 4 | 26 | | | | | |
| a2 No escapable | 5 | 39 | | | | | |
| a2 No escapable | 6 | 38 | | | | | |
| a2 No escapable | 7 | 23 | | | | | |
| a2 No escapable | 8 | 28 | | | | | |
| SC (Suma de Cuadrados) | | | | | | | |
| gl (grados de libertad) | | | | | | | |
| MC (Media Cuadrática) | | | | | | | |
| | | | | SCtotal | SCentre | | SCerror |

Figura 23. Inicio del desarrollo de la ecuación estructural y ANOVA

Llegados a este punto, solo queda por determinar si el ajuste del modelo a los datos es lo suficientemente considerable como para concluir que es estadísticamente significativo (se trata de ejecutar la decisión estadística) junto cono estimar el valor del tamaño del efecto y su intervalo de confianza.

En el ejemplo se pretende comprobar la hipótesis que mantiene que la situación de desamparo aprendido provoca déficits depresivos operacionalizados en conductas de frustración y desinterés. Si los resultados obtenidos confirman o no esta hipótesis se comprueba calculando la probabilidad de obtener el tamaño del efecto detectado en el experimento cuando la hipótesis teórica que sustenta el modelo es falsa, es decir, no hay relación entre la variable independiente (shock) y la variable dependiente (tiempo) (planteamiento de la hipótesis nula). Para esto se necesita conocer el número de resultados que pueden producirse en un determinado experimento (el universo de resultados) y la probabilidad asociada con el ejemplo, para poder determinar si el resultado obtenido es muy improbable o extraño (es estadísticamente significativo) o, por el contrario, está dentro del intervalo de confianza (Arnau, 1981).

A continuación se detalla un resumen de todos los pasos implicados en el proceso de contraste estadístico.

En resumen, la prueba de la hipótesis experimental se inicia partiendo del supuesto de que no existe relación entre las variables independientes y las dependientes, a este enunciado se le denomina hipótesis nula, explicando las diferencias que pudiesen aparecer por el efecto del azar atribuido al error de muestreo. Es decir, si el experimento se repitiera extrayendo las infinitas muestras de la variable Y en las distintas condiciones del factor A y se calculará la media en cada una de las a condiciones de la variable, la esperanza matemática de cada $\bar{Y}_{a(i)}$ sería igual al valor de la media poblacional de Y (μ). El supuesto de la distribución muestral que se comprueba se denomina prueba de significación de la hipótesis nula.

En vez de recoger todas las posibles muestras del mismo experimento para determinar los parámetros de la población muestral con exactitud, se asume un determinado límite de error que se denomina alfa, α . El valor de α o error de Tipo I limita la probabilidad de equivocarse al rechazar la hipótesis nula (que se considera que realmente es cierta). Como la prueba de la hipótesis se realiza conociendo la distribución muestral del estadístico F entonces se conoce exactamente la probabilidad acumulada en cada valor del mismo, siendo la probabilidad de error de Tipo I asociada con cada valor de este estadístico (valor p) la probabilidad complementaria de dicha probabilidad acumulada.

El nivel de significación está establecido por consenso en la comunidad científica en el 5% si no se indica otro valor en la fase de planificación del estudio y, de ahí, que no sea necesario especificar dicho valor si el investigador va a ejecutar sus contrastes de hipótesis con el máximo de error de Tipo I que se considera apropiado en las Ciencias del Comportamiento. La probabilidad complementaria se le denomina nivel de confianza ($1 - \alpha$) y determina cuál será la probabilidad de que el investigador acierte en su decisión cuando la hipótesis nula sea cierta dado que no existe relación entre las variables independientes y las dependientes del modelo.

Una vez se determina el tamaño del efecto experimental del estudio (se conoce el valor de la prueba estadística ejecutada) se calcula su probabilidad bajo el supuesto de la hipótesis nula. Se puede concluir que el resultado es

estadísticamente significativo y rechazar la hipótesis nula ($p \leq \alpha$), o por el contrario mantenerla si la probabilidad asociada con el tamaño del efecto detectado supera el punto α preestablecido ($p > \alpha$). El estadístico que se emplea para realizar dicha comprobación es la razón F que es la razón entre la variación de los datos observada entre los distintos niveles de un tratamiento (Media Cuadrática del Efecto: Sumas de Cuadrados del efecto o del factor / sus grados de libertad) respecto de la suma de cuadrados del término del error (Media Cuadrática del Error: Sumas de Cuadrados del error / sus grados de libertad).

Como cada una de estas sumas de cuadrados se estiman con un número distinto de observaciones independientes, se corrige estas diferencias dividiendo cada suma de cuadrados por sus correspondientes grados de libertad (g).

Los grados de libertad indican el número de observaciones cuyos valores son libres de variar, o en otras palabras, el número de observaciones independientes de una fuente de variabilidad menos el número de parámetros estimados al computar dicha variación.

Grados de libertad totales

Los grados de libertad de la suma de cuadrados total son $N - 1$. Es decir, la suma cuadrática de la variación total, $\sum(Y - M)^2$, implica N observaciones independientes y se pierde un grado de libertad cuando se estima el parámetro desconocido μ con la media total de las observaciones, M . Entonces como la suma de las desviaciones se realiza de cada puntuación respecto a la media general, $N - 1$ observaciones pueden tomar cualquier valor (tienen grados de libertad), ya que solamente una de ellas está determinada (no tiene grados de libertad para variar) para poder obtener el valor concreto de la media total de los datos.

Grados de libertad del efecto

Los grados de libertad de la suma de cuadrados de la fuente de tratamiento es igual $a - 1$, pues la puntuación predicha \hat{Y} en $\sum(\hat{Y} - M)^2$ contiene a parámetros a estimar, $\hat{b}_0, \hat{b}_1 \dots \hat{b}_k$ y se sabe que el sumatorio de las desviaciones $\hat{Y} - M$ tiene que ser igual a cero, entonces únicamente $a - 1$ de los parámetros a estimar son libres de variar.

Grados de libertad del error

Los grados de libertad de la suma de cuadrados del error o residual son iguales a $N - a$, ya que la suma de cuadrados del error, $\sum(Y - M)^2$, implica N observaciones independientes donde la puntuación predicha \hat{Y} supone estimar a parámetros independientes dentro de cada condición experimental.

Los resultados del supuesto señalan que con 1 grado de libertad de la fuente del tratamiento y 6 de la del error se obtiene que:

$$F(1, 6) = \frac{\frac{392}{1}}{\frac{356}{6}} = \frac{392}{59.334} = 6.607$$

Concluyendo, se rechaza el modelo de la hipótesis de nulidad, ya que el valor de $F(1, 6) = 6.607$ tiene una probabilidad menor del riesgo de error de Tipo I fijado a priori por el investigador (en concreto es de $p = .042$) o lo que es lo mismo el valor empírico de la razón F es mayor que el valor de la distribución teórica que le corresponde con 1 y 6 grados de libertad para el término del efecto y del error respectivamente ($F_{\text{teórica}}(1, 6) = 5.987$). El valor de la denominada teórica o de tablas F_t ($\alpha = .05, 1, 6$) es de 5.987, y únicamente indica que el 95% de los valores que pueden obtenerse de F en la muestra del experimento, cuando la hipótesis nula sea cierta, serán menores o iguales que 5.987.

Redacción de resultados

La redacción de los resultados del *diseño entre-grupos univariados univariado* en un informe de investigación siguiendo el formato del Manual del APA sería, por ejemplo, la siguiente:

Los resultados del estudio del efecto de la indefensión aprendida sobre la sintomatología depresiva mediante un diseño entre-grupos unifactorial univariado con dos grupos (grupo 1 = shock eléctrico escapable, grupo 2 = shock eléctrico no escapable) señalan que las ratas que son sometidas a un situación de indefensión (reciben shock eléctrico no escapable o independiente de su conducta) tienen una puntuación media más alta en depresión (Media = 32, $DT = 7.79$, $n = 4$) que las ratas que sí tienen control de la situación del shock y, por lo tanto, no desarrollan la situación de indefensión aprendida (Media = 18, $DT = 7.62$, $n = 4$), siendo la diferencia entre las medias de los dos

grupos estadísticamente significativa con un tamaño del efecto grande, $F(1, 6) = 6.61$, $p = 0.042$, $\eta^2 = .52$. Por lo tanto, observando las puntuaciones medias de las condiciones experimentales, las ratas del grupo de shock no escapable recorrieron el laberinto con mayor lentitud (media = 32) que las ratas que fueron sometidas a shock escapable (media = 18).

Por lo tanto, cuando se redactan los resultados del contraste de hipótesis es fundamental dar toda la información estadística que rodea al proceso de contraste de hipótesis estadística como el tipo de diseño y las variables implicadas en el análisis que se ejecuta en la ecuación estructural, los descriptivos, al menos, de media y su desviación típica, y número de observaciones y el valor de la razón F con sus grados de libertad, junto con su valor p exacto y la información del tamaño del efecto (en este caso se ha interpretado la proporción de varianza explicada conocida como eta cuadrado η^2 cuyo calculo es suma de cuadrados del efecto dividido por la suma de cuadrados total) y su intervalo de confianza (en la redacción no se ha incluido la información del intervalo de confianza de eta cuadrado).

En la siguiente dirección (<https://www.gigacalculator.com/>) se pueden llevar a cabo de forma rápida la estimación de los valores de media y desviación típica para que se incluyan en el apartado de redacción de los resultados:

<https://www.gigacalculator.com/calculators/standard-deviation-calculator.php>

Además, en la dirección <https://www.gigacalculator.com/calculators/statistics/> hay un amplio abanico de herramientas para realizar cálculos estadísticos.

Aunque la prueba de la hipótesis se ha formulado para comprobar cierto orden en las medias de las condiciones de las variables independientes, la prueba de la hipótesis del estadístico F compara varianzas; de ahí que el nombre que reciba esta prueba sea precisamente el de análisis de la varianza. Como ya se ha comentado, la prueba estadística es completamente válida en la medida que se cumplan dos supuestos fundamentales de las pruebas paramétricas: que la variable dependiente se distribuya normalmente y que las distintas observaciones sean independientes y extraídas aleatoriamente, no existiendo entre las condiciones experimentales ninguna otra diferencia que la manipulación de la variable independiente o la relación predicha entre las variables.

En definitiva, el planteamiento teórico que el investigador desarrolla en su hipótesis determinará la selección de un diseño de investigación concreto cuyo plan de investigación debe seguir la pauta de *maximización* de la varianza sistemática primaria, *minimización* de la varianza del error y *control* de la varianza sistemática secundaria (principio MAX-MIN-CON) Las características de la investigación y especialmente la posibilidad o no de la asignación aleatoria a las condiciones de tratamiento determinará la modalidad o metodología de investigación.

La proporción de varianza explicada: R^2 , η^2 y η^2 parcial

R^2 y η^2

En el análisis de regresión el tamaño del efecto se pueden estimar con el coeficiente de determinación (R^2). El coeficiente de determinación cuantifica la proporción de varianza de la variable respuesta o variable dependiente que es explicada por el efecto de la variable predictora o variable independiente. Su valor es el mismo que el del tamaño del efecto eta cuadrado (η^2) que se estima en el Análisis de la Varianza (ANOVA). La principal ventaja de estos índices (proporciones) es su fácil interpretación ya que se puede multiplicar por 100 y hablar en términos de porcentaje de varianza explicada por el efecto de la variable independiente. Por ejemplo, si el valor de η^2 es .15 entonces el 15% de las diferencias encontradas entre los dos grupos se atribuye al efecto de la intervención o tratamiento. En términos de Cohen se trataría de un tamaño del efecto grande. Además, $1 - \eta^2$ es la proporción de varianza no explicada por el efecto del tratamiento y que se atribuye por lo tanto a la fuente de varianza del error. En el ejemplo anterior, el 85% de la varianza detectada entre las puntuaciones no está explicada por el efecto de la variable independiente y se atribuye a la varianza del error.

Los índices vinculados al análisis de la varianza (ANOVA) como proporción de varianza eta cuadrado (η^2) y eta cuadrado parcial (η_p^2) se pueden obtener a partir de las sumas de cuadrados de la tabla del ANOVA.

El cálculo de eta cuadrado (η^2) o proporción de varianza explicada por el efecto del tratamiento es:

$$R^2 = \eta^2 = \frac{\text{Sumade Cuadrados}_{\text{TRATAMIENTO o EFECTO}}}{\text{Sumade Cuadrados}_{\text{TOTAL}}}$$

Diseños factoriales: η^2 y η^2_{parcial}

En los diseños donde se incluye la estimación del efecto de más de una variable independiente o factor (“diseños factoriales”), el tamaño del efecto puede estimarse para cada fuente de varianza de forma total (η^2) o de forma parcial (η_p^2). A continuación se detallan las características de cada estadístico utilizando los datos del ANOVA factorial 3 x 2 que se detallan en la figura 24.

Pruebas de efectos inter-sujetos

Variable dependiente: Y

| Origen | Tipo III de suma de cuadrados | gl | Cuadrático promedio | F | Sig. | Eta parcial al cuadrado |
|------------------|-------------------------------|----|---------------------|---------|------|-------------------------|
| Modelo corregido | 42,000 ^a | 5 | 8,400 | 6,300 | ,004 | ,724 |
| Interceptación | 450,000 | 1 | 450,000 | 337,500 | ,000 | ,966 |
| A | 12,000 | 2 | 6,000 | 4,500 | ,035 | ,429 |
| B | 18,000 | 1 | 18,000 | 13,500 | ,003 | ,529 |
| A * B | 12,000 | 2 | 6,000 | 4,500 | ,035 | ,429 |
| Error | 16,000 | 12 | 1,333 | | | |
| Total | 508,000 | 18 | | | | |
| Total corregido | 58,000 | 17 | | | | |

a. R al cuadrado = ,724 (R al cuadrado ajustada = ,609)

Figura 24. Inicio del desarrollo de la ecuación estructural y ANOVA

Eta Cuadrado: η^2

Cuando se trata de estimar el tamaño del efecto de η^2 de forma total entonces hay que dividir la Suma de Cuadrados del Efecto por la Suma del Cuadrados Total. Si se trata de un diseño con dos factores o dos variables independientes (A x B) entonces el tamaño del efecto se estima para los dos efectos principales (A y B) y para el efecto de interacción A x B. Por lo tanto, habrá que estimar tres tamaños del efecto:

$$\eta_{\text{EFECTO A}}^2 = \frac{\text{Sumade Cuadrados}_{\text{EFECTO A}}}{\text{Sumade Cuadrados}_{\text{TOTAL}}}$$

$$\eta_{\text{EFECTO B}}^2 = \frac{\text{Sumade Cuadrados}_{\text{EFECTO B}}}{\text{Sumade Cuadrados}_{\text{TOTAL}}}$$

$$\eta_{EFECTOAB}^2 = \frac{SumadeCuadrados_{EFECTOAB}}{SumadeCuadrados_{TOTAL}}$$

Los resultados de eta cuadrado calculado de forma total para la tabla del ANOVA 2 x 3 presentada en la figura 24 son los siguientes:

$$\eta_{EFECTO A}^2 = \frac{12}{58} = .207$$

$$\eta_{EFECTO B}^2 = \frac{18}{58} = .310$$

$$\eta_{EFECTO AB}^2 = \frac{12}{58} = .207$$

Eta Cuadrado parcial: η_p^2

El estadístico de eta cuadrado parcial (η_p^2) es la proporción de varianza explicada por el efecto (efecto de A, efecto de B o efecto de interacción AB) más la del error que se puede atribuir a dicho efecto o fuente de varianza. Cuando se trata de estimar el tamaño del efecto de η^2 parcial entonces hay que dividir la Suma de Cuadrado del Efecto a estimar por la Suma del Cuadrados del Efecto que se está estimado más la Suma de Cuadrados del Error. Es decir, el denominador de la fórmula incluye la suma de cuadrados del efecto que se está considerando más la suma de cuadrados del error. De nuevo, si se utiliza un diseño factorial se pueden estimar tres tamaños del efecto, dos para los efectos principales (A y B) y uno para el efecto de interacción (AB).

$$\eta_p^2 = \frac{SumadeCuadrados_{TRATAMIENTO EFECTO}}{SumadeCuadrados_{TRATAMIENTO EFECTO} + SumadeCuadrados_{ERROR}}$$

Por lo tanto, la estimación del valor de η_p^2 para cada una de las tres fuentes de varianza de un diseño factorial A x B es la siguiente:

$$\eta_{pEFECTO A}^2 = \frac{SumadeCuadrados_{EFECTO A}}{SumadeCuadrados_{EFECTO A} + SumadeCuadrados_{ERROR}}$$

$$\eta_{pEFECTO B}^2 = \frac{SumadeCuadrados_{EFECTO B}}{SumadeCuadrados_{EFECTO B} + SumadeCuadrados_{ERROR}}$$

$$\eta_{pEFECTO AB}^2 = \frac{SumadeCuadrados_{EFECTO AB}}{SumadeCuadrados_{EFECTO AB} + SumadeCuadrados_{ERROR}}$$

Los resultados de eta cuadrado calculado de forma parcial para la tabla del ANOVA 2 x 3 presentada en la figura 24 se presentan a continuación. Estos valores son los que se detallan en el SPSS dado que dicho programa estadístico únicamente calcula la η_p^2 .

$$\eta_{EFECTO A}^2 = \frac{12}{12+16} = .429$$

$$\eta_{EFECTO B}^2 = \frac{18}{18+16} = .529$$

$$\eta_{EFECTO AB}^2 = \frac{12}{12+16} = .429$$

Cuando se trabaja con diseños factoriales, Pierce, Block y Aguinis (2004) subrayan las precauciones que los investigadores deben adoptar cuando interpretan el valor de eta cuadrado parcial. En este caso el valor de eta cuadrado parcial es mayor para una determinada fuente de varianza que el valor de eta cuadrado. Por lo tanto, la η_p^2 sobreestima el tamaño del efecto.

En definitiva, solamente cuando el diseño tiene una única variable independiente o factor los valores de eta cuadrado y eta cuadrado parcial coinciden. Además, cuando se suman los valores de eta cuadrado parcial de las fuentes de varianza de un diseño factorial se puede obtener un valor superior a uno. Con los datos de la figura 24 se observa que la suma de los valores de η_p^2 es igual a 1.387. Sin embargo, la suma de los valores de eta cuadrado de un determinado diseño factorial nunca puede ser superior a uno ya que se computan con el mismo término de error. Por ello, es importante redactar en el informe o artículo qué estadístico del tamaño del efecto se ha calculado, recordando que el SPSS siempre computa el valor de η_p^2 , sobreestimando el tamaño del efecto. Con los datos de la figura 24 se observa que la suma de los valores de η^2 es igual a .724, y si se suma la varianza atribuida al error o varianza no explicada por los efectos del tratamiento ($1 - \eta^2 = 0.276$) entonces la suma total siempre será igual a 1.

Comprobación de hipótesis específicas (diseño entre grupos $A < 2$)

Hasta ahora se ha presentado la prueba paramétrica para contrastar hipótesis de dos grupos independientes (diseño entre-grupos unifactorial $A = 2$ univariado) mediante el Análisis de la Varianza (ANOVA). Ahora nos vamos a detener en el contraste de hipótesis específicas. Se van a presentar diferentes pruebas estadísticas y se van a utilizar dos módulos del SPSS donde se pueden realizar las pruebas de hipótesis específicas en la ventana de 'Analizar' y luego se opta por el 1) módulo de 'Comparar medias' o por el 2) módulo de 'Modelo Lineal General'.

ANALIZAR → Comparar medias

o

ANALIZAR → Modelo Lineal General

El SPSS tiene un grupo de análisis dentro del menú Analizar → Comparar Medias donde se pueden ejecutar las técnicas de contraste de hipótesis de prueba t para una muestra, prueba t para muestras independientes, prueba t para muestras relacionadas y ANOVA de un factor.

Nos vamos a detener en el último apartado: ANOVA de un factor (figura 25). Dentro de este apartado se puede ejecutar un diseño de $A = 2$ y de $A > 2$. Es decir, se pueden ejecutar los diseños unifactoriales con un factor de 2 condiciones o con un factor que tenga más de dos condiciones (por ejemplo $A = 3$).

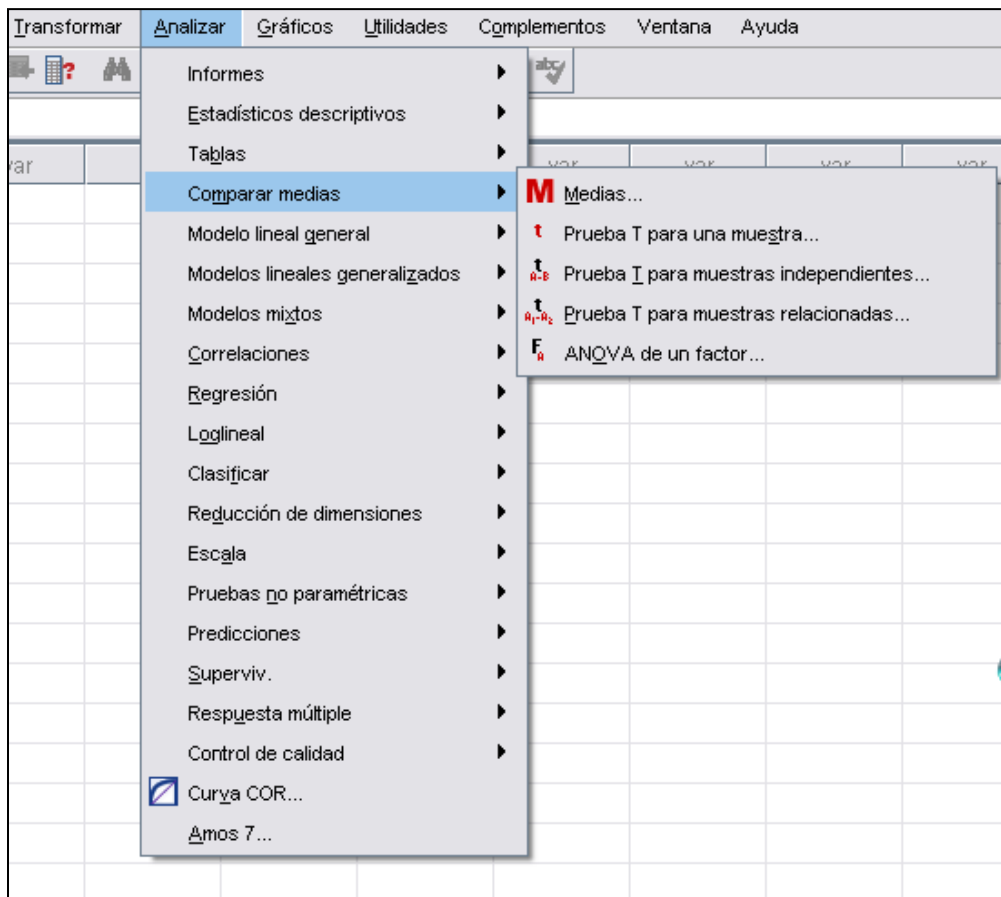


Figura 25. SPSS: Analizar-Comparar medias-ANOVA de un factor

Cuando el diseño tiene un factor o variable independiente con más de dos condiciones será necesario recurrir a la opción de ‘post-hoc’ (a posteriori) para poder descubrir entre qué pares de medias se encuentran las diferencias (figura 26). Y para ello es necesario que el profesional (investigador o investigadora) decida qué prueba debe ejecutar a posteriori, teniendo en cuenta si se asume que las varianzas de los grupos son homogéneas (“homocedasticidad de varianzas”) o no lo son (“heterocedasticidad de varianzas”).

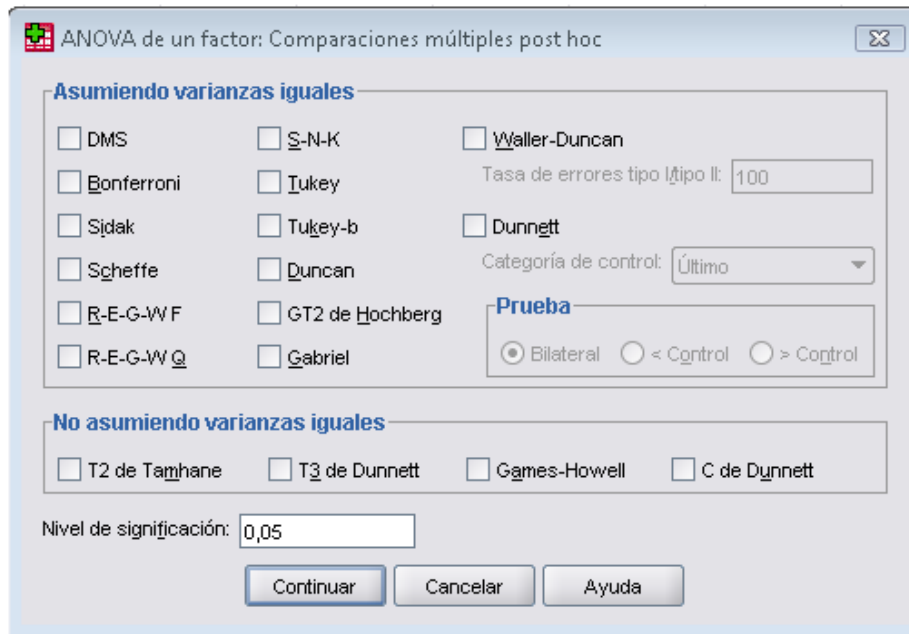


Figura 26. SPSS: Analizar-Comparar medias-ANOVA de un factor

Por lo tanto, cuando la variable independiente tiene más de 2 condiciones, hay que analizar entre qué medias se producen las diferencias estadísticamente significativas y en qué sentido.

Supongamos que una investigación tiene un diseño entre-grupos unifactorial univariado, $A = 3$ ($N = 9$, $n_a = 3$), y tiene los siguientes datos:

$$a_1 = 12, 8, 10$$

$$a_2 = 5, 7, 6$$

$$a_3 = 14, 13, 15$$

El investigador desea comprobar si existen diferencias estadísticamente significativas entre esos tres grupos y también entre qué par de medias se encuentran las diferencias que son estadísticamente significativas.

Se lleva a cabo la introducción de datos para este tipo de diseño en el SPSS y, posteriormente, se ejecutan las instrucciones del modelo de diseño que se desea analizar (figura 27).

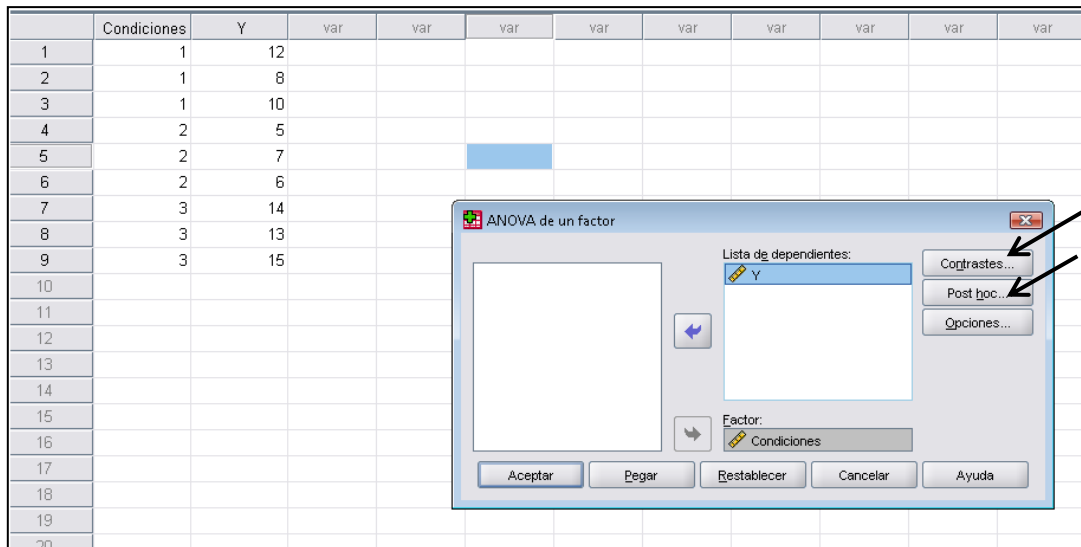


Figura 27. SPSS: Introducción de datos para un diseño entre-grupos y selección de la variable dependiente e independiente del diseño

Si se ejecuta el ‘ANOVA de un factor’ y en ‘Opciones’ se selecciona ‘descriptivos’ y ‘prueba de homogeneidad’ de las varianzas, el resultado señala la siguiente relación entre las medias de los grupos mediante la razón F del Análisis de la Varianza (figura 28). El lector o lectora debe responder a la pregunta de si *¿existe alguna diferencia estadísticamente significativa* entre las medias de las tres condiciones?

A continuación se describe la información que proporciona el ANOVA y las pruebas de hipótesis específicas.

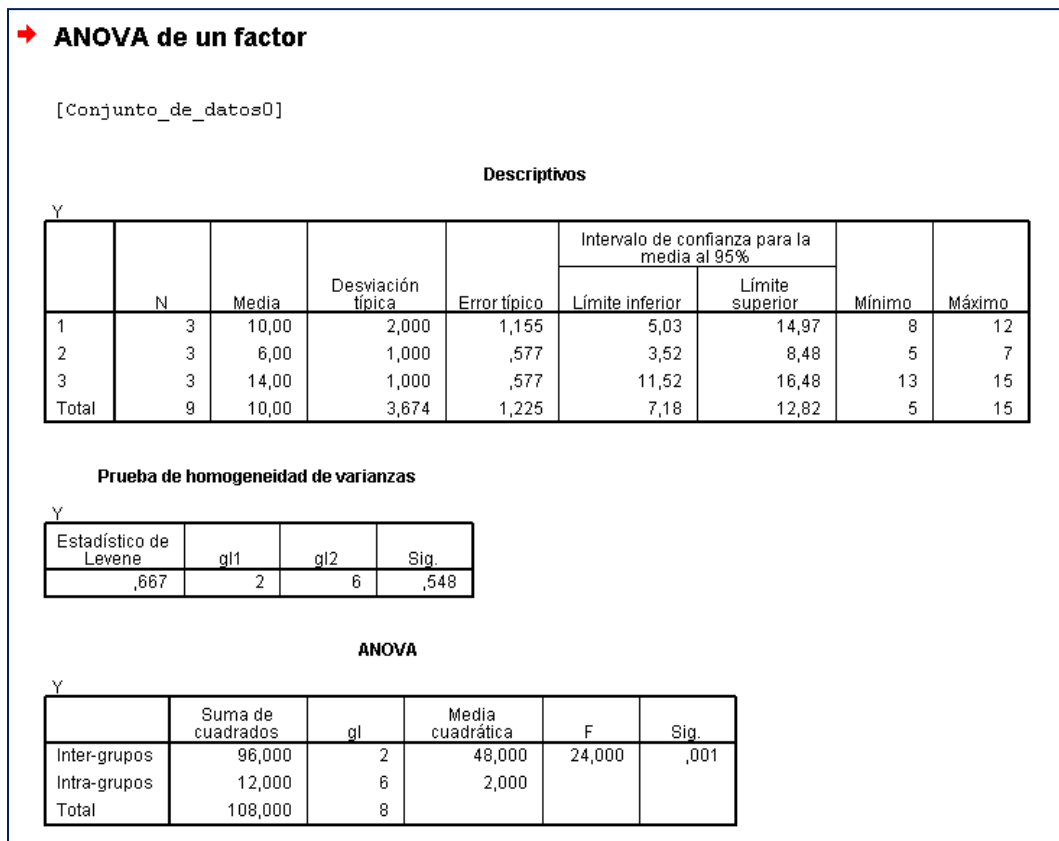


Figura 28. SPSS: resultados del ANOVA y homogeneidad de varianzas

Los resultados de la prueba de homogeneidad de varianzas de Levene señalan que las varianzas de las puntuaciones de las tres condiciones son homogéneas (Levene_(2, 6) = 0.667, $p = .548$). Por lo tanto se cumple el supuesto de homogeneidad de las varianzas.

Los resultados del modelo de ANOVA de un diseño entre-sujetos unifactorial A = 3 señalan que existen diferencias estadísticamente significativas entre los tres grupos ($F_{(2, 6)} = 24$, $p = .001$).

Pero, cuando $A > 2$, la pregunta es ¿dónde se encuentran las diferencias que son estadísticamente significativas? ¿Es entre las medias de $a_1 - a_2$? ¿Es entre las medias de $a_1 - a_3$? O quizás ¿es entre las medias de $a_3 - a_2$? Las diferencias siempre se valoran en términos de valores absolutos.

El resultado del ANOVA no permite dar respuesta a esas comparaciones específicas entre los tres grupos. Cuando únicamente se trabaja con dos medias ($A = 2$) estaba claro que si el resultado del ANOVA señalaba que habían diferencias estadísticamente entre los grupos se trataba de la única diferencia posible: $a_1 - a_2$.

Pero ahora hay tres posibles diferencias de medias. ¿Cuál de ellas es estadísticamente significativa? ¿Son todas? ¿Es una? ¿Son dos? En las pruebas post-hoc de comparación de medias (también conocidas como pruebas a posteriori) se encuentra la solución a esos interrogantes.

Antes de continuar conviene recordar cómo se desarrolla la ecuación estructural para el modelo planteado en el estudio con $A=3$:

$$Y = M + A + E$$

Las puntuaciones en la variable dependiente Y son igual a:

La constante (media general) + el efecto del tratamiento (Efecto $A = \mu_a - \mu$) + Error (Error = $Y - M - \text{Efectos}$, es decir, $Y - M - A$ en este tipo de diseño).

Además, se puede calcular:

- Los efectos estimados para el diseño planteado son $\alpha_1 = 0$ y $\alpha_2 = -4$ (luego $\alpha_3 = 4$ ya que la suma de los efectos es 0)
- Y los errores son en $a_1 = 2$ y -2 , en $a_2 = -1$ y 1 y en $a_3 = 0$ y -1 .

Como ejercicio, los lectores y lectoras pueden completar la tabla de efectos y los errores y desarrollar el modelo de diseño propuesto mediante la descomposición de la ecuación estructural. Se debe comprobar que la suma de los efectos del tratamiento es cero así como la del error. Elevar al cuadrado y sumar (Suma de Cuadrados, SC). Dividir cada suma de cuadrados por sus grados de libertad (gl) (Medias Cuadráticas, MC) y ejecutar la prueba F del Análisis de la Varianza ($F = \text{MC}_{\text{efecto}} / \text{MC}_{\text{error}}$). Después, se debe utilizar las tablas de la razón F y buscar el valor teórico que corresponde a $F(.05, 2, 6)$. A continuación, se comparan los dos valores de F y si el valor de la F empírica del estudio es mayor o igual que el valor de la F teórica de las tablas entonces se puede rechazar la hipótesis nula (es decir, se rechaza la hipótesis nula cuando $F(2, 6) \geq F(.05, 2, 6)$). Cuando se utilizan las tablas sólo se puede saber si el valor p de probabilidad del resultado de la prueba estadística es superior o menor al valor de alfa. Por eso, cuando se utilizan las tablas no se puede dar el valor exacto de probabilidad del resultado, pero en nuestro ejemplo sí se puede saber que $p < .05$ pues la F empírica es mayor que la F teórica. Por lo tanto, se puede completar el proceso de decisión estadística tal y como se efectúa en el contraste de hipótesis.

Cuando se trabaja con un programa estadístico (como por ejemplo SPSS; JASP, JAMOVI, SAS...) ya no hace falta recurrir a las tablas de la distribución del estadístico. El mismo programa añade el valor exacto de probabilidad (valor p , a veces se señala como significación) junto al resultado de la prueba estadística. El investigador o investigadora tiene que completar el proceso de contraste estadístico comparando el valor de p de probabilidad del resultado obtenido (o más extremo) con la prueba estadística con el valor del alfa fijado a priori en su investigación. Si el valor de $p \leq \alpha$ entonces se puede rechazar la hipótesis nula. Siempre que se trabaja con un programa estadístico hay que redactar los valores p de probabilidad exactos incluso cuando las decisiones suponen mantener la hipótesis nula ($p > \alpha$). No es nada recomendable situar 'ns' para indicar un resultado no estadísticamente significativo. Recordemos que el valor p de probabilidad depende del valor del efecto y del tamaño de la muestra. Así, manteniendo constante el efecto, si se aumenta la muestra disminuye el valor de probabilidad. Del mismo modo, si se disminuye el tamaño de la muestra entonces aumenta el valor p de probabilidad.

Por lo tanto, cuando el Análisis de la Varianza (ANOVA) ofrece un resultado estadísticamente significativo está indicando que al menos un grupo difiere de los otros grupos. EL ANOVA es un 'test omnibus', es decir, no informa del patrón de diferencias entre las medias, sólo facilita que hay (o no hay) alguna diferencia estadísticamente significativa entre las medias de los grupos, pero sin señalar dónde está la diferencia o las diferencias. Es decir, el ANOVA solamente detecta la presencia o ausencia de un efecto global de la variable independiente sobre la variable dependiente. Pero el investigador o investigadora desea conocer dónde se encuentran las diferencias estadísticamente significativas que detectó de forma global el ANOVA. Para analizar el patrón de diferencias entre las medias es necesario ejecutar pruebas de hipótesis específicas y las más utilizadas implican comparar pares de medias, de ahí el nombre de comparaciones por parejas ('pairwise comparisons'). En el SPSS se conocen como 'pruebas 'post hoc'.

Siguiendo con el ejercicio, una vez que se ha comprobado que existe un efecto global en el Análisis de la Varianza tal y como se ha detallado en la salida del SPSS ($F_{(2, 6)} = 24, p = .001$), se puede abordar entre qué pares de medias se encuentran

las diferencias estadísticamente significativas. Para ello, se puede elaborar la tabla de diferencias de medias entre los grupos o condiciones (figura 29).

| | | |
|------------------|------------------|-----------------|
| Medias | a1 Media = 10 | a2 Media = 6 |
| a2 Media = 6 | 4 | -- |
| a3 Media = 14 | 4 | 8 |

Figura 29. Diferencias de medias para un diseño A = 3

Con un diseño A = 3 existen 3 diferencias de medias entre los 3 grupos. Para conocer cuántas diferencias de medias simples existente en un diseño se puede aplicar la siguiente fórmula:

$$C = \frac{a(a-1)}{2}$$

Donde C es el número de diferencias de medias simples entre los grupos no redundantes y a es el número de condiciones que tiene el modelo de diseño.

Entonces, ¿cuál de las tres diferencias de medias es estadísticamente significativa? Para responder a dicha cuestión se necesita ejecutar una prueba de contraste de hipótesis específicas para comparar las diferencias de medias entre las condiciones experimentales.

Una solución que se podría pensar es ejecutar pruebas *t* de Student dos a dos o ANOVAs para cada par de medias ejecutando en este caso 3 ANOVAs. Pero esta alternativa no se considera válida ya que está sujeta a una grave deficiencia metodológica como es aumentar la probabilidad del error de Tipo I. Es decir, la tasa de error de Tipo I por experimento deja de ser la planteada como alfa cuando se planificó el estudio. Para poder realizar esos contrastes estadísticos correctamente es necesario controlar el alfa por comparación. Veamos en primer lugar qué es la Tasa de Error de Tipo I y en segundo lugar cómo controlar dicha tasa de error mediante los procedimientos de pruebas de hipótesis específicas.

Tasa de error de tipo I

La Tasa de Error de Tipo I o alfa por experimento (α_{PE}) es:

$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^C$$

Donde alfa por comparación, α_{PC} , es la probabilidad de cometer un error de tipo I en 'una comparación' y el alfa por experimento, α_{PE} , es la probabilidad de cometer al menos un error de Tipo I en un 'conjunto de comparaciones' (C).

Supongamos que en un estudio se van a realizar 4 comparaciones o pruebas de contraste de hipótesis. Si todas las hipótesis nulas fueran ciertas y $\alpha_{PC} = .05$ entonces la probabilidad de cometer al menos un Error de Tipo I es:

$$\alpha_{PE} = 1 - (1 - .05)^4 = .1855$$

Ese valor de .1855 se aparta mucho del valor de alfa fijado a priori por el investigador en .05. Es decir, se tendría una probabilidad de error de Tipo I de .1855, que es totalmente inadmisibles. La probabilidad de rechazar una hipótesis nula siendo cierta es de .1855 (18.5%). ¿Qué ha pasado? La prueba de contraste de hipótesis se ha hecho más liberal (es decir, se rechaza la hipótesis nula con mayor facilidad), aumentando de este modo el error de Tipo I. ¿Cómo se puede corregir ese sesgo que amenaza la validez de conclusión estadística de los resultados del estudio? La opción es ejecutar un procedimiento que haga a la prueba estadística más conservadora para equilibrar ese aumento del error de Tipo I.

Todos los procedimientos de contraste de hipótesis específicas se basan en hacer el contraste más conservador. Es decir, se reduce el α_{PC} (alfa por comparación) para poder controlar el α_{PE} (alfa por experimento) que no debe superar el valor de .05. En definitiva, la prueba se hace más conservadora.

En este punto el investigador o investigadora debe tomar una decisión de nuevo: tiene que elegir la prueba de contraste de hipótesis específicas que controle correctamente la tasa de error de Tipo I y además que la potencia estadística sea máxima (menor error de Tipo II). Es decir, en un mismo diseño se podría optar por varias pruebas de contraste de hipótesis específicas y el investigador o investigadora debe seleccionar la más adecuada para que la validez de conclusión

estadística sea la más óptima. La más adecuada es aquella que controla la tasa de error de tipo I y fija el alfa por experimento en .05 y al mismo tiempo es la prueba que facilita la mayor potencia estadística para de este modo reducir el error de Tipo I.

Para tomar la decisión el investigador o investigadora debe considerar los siguientes aspectos:

1. El número de comparaciones (C) que la hipótesis plantea: pueden ser comparaciones exhaustivas (*a posteriori*) o pueden ser comparaciones planificadas (*a priori*).
2. Si las hipótesis experimentales son simples (plantea diferencias *entre pares de medias*) o complejas (plantea alguna diferencia entre medias que implica *un promedio de medias*).

Comparaciones exhaustivas o a posteriori. El contraste de hipótesis específicas es exhaustivo cuando se realizan todas las comparaciones posibles entre los padres de medias que tiene el diseño de investigación. Por ejemplo, si $A = 3$ entonces el número de todos los pares posibles de diferencias de medias es igual a 3. Si se analizan las tres diferencias simples de medias entonces se considera que se han realizado comparaciones exhaustivas.

Comparaciones planificadas o a priori. Cuando el número de comparaciones que hay que contrastar es más reducido (no se realizan de forma exhaustiva todas las comparaciones simples), el contraste se denomina contraste planificado o contraste a priori. Por ejemplo, si sólo se desean comparar $a_1 - a_2$ y $a_1 - a_3$ entonces ya no es exhaustivo sino planificado dado que por cuestiones teóricas no interesa analizar la comparación entre a_2 y a_3 .

Para poder plantear contrastes a priori es necesario fundamentar esas opciones de contraste en unas hipótesis teóricas que den sentido a la elección de los contrastes o análisis que se quieren realizar. Por ejemplo, en un diseño con un grupo de control y dos grupos de tratamiento ($A = 3$) podría ser interesantes comparar el grupo de control con uno de los grupos de tratamiento (contraste uno) y el grupo de control con el otro grupo de tratamiento (contraste dos). En este caso el investigador ha planteado dos contrastes o pruebas consideradas a priori, es

decir, su planteamiento es previo a cualquier tipo de resultado del estudio. Los contrastes se plantean antes de ejecutar cualquier análisis y necesitan de una teoría que los fundamente.

Las **hipótesis son simples** cuando se plantean diferencias entre pares de medias simples, por ejemplo la diferencia entre las medias de los grupos $a_1 - a_2$. Se considera que las **hipótesis son complejas** cuando la diferencia de medias implica un promedio de medias, por ejemplo se desea comparar la media de a_1 frente a la media compleja de dos grupos. Por ejemplo una media compleja sería $(a_2 + a_3) / 2$. Es decir, en el cómputo de la media han intervenido las medias de dos grupos. Al final se contrastan dos medias pero al menos una de ellas es compleja. Otro ejemplo de hipótesis complejas podría ser la comparación siguiente: la media de $(a_2 + a_3) / 2$ respecto a la media $(a_1 + a_3) / 2$.

Pruebas de contraste de hipótesis específicas

Existen diferentes pruebas de contraste de hipótesis específicas. En la figura 30 se representan las situaciones en las que se podrían aplicar las *pruebas de hipótesis específicas* más utilizadas:

- **Bonferroni:** hipótesis planificadas, simples y complejas.
- **Dunnett:** hipótesis planificadas y simples.
- **DHS Tukey:** hipótesis exhaustivas y simples.
- **Scheffé:** hipótesis planificadas, exhaustivas, simples y complejas.

| | Planificada | Exhaustiva | Simple | Compleja |
|------------|-------------|------------|--------|----------|
| Bonferroni | X | | X | x |
| Dunnett | x | | X | |
| DHS Tukey | | x | X | |
| Scheffé | x | x | x | x |

Figura 30. Diferentes pruebas de contraste de hipótesis específicas

Como se observa, hay situaciones donde se pueden aplicar varias pruebas de hipótesis específicas, pero dependiendo de la situación hay unas pruebas que tienen más potencia estadística (menor error de Tipo II). Recordar que el

investigador o investigadora tiene que seleccionar aquella prueba que además de controlar el error de Tipo I también sea la que tiene más potencia estadística para detectar el efecto (menor error de Tipo II).

A continuación se detallan las situaciones de investigación donde es más apropiada una prueba estadística u otra, dado que controlan la probabilidad del error de Tipo I en el nivel de alfa planteado por el investigador a priori y tienen el menor error de Tipo II.

Procedimiento DHS (Honestly Significant Difference) de Tukey

El procedimiento DHS de Tukey es el más potente cuando en el diseño se ejecutan:

- 1) todas las comparaciones posibles entre las medias (exhaustivo) y además
- 2) son comparaciones simples

Es decir, es la prueba más adecuada cuando se comparan dos medias simples cada vez y se ejecutan todas las posibles comparaciones simples del diseño.

El procedimiento calcula el denominado 'rango crítico' (RC) o valor teórico de la prueba y se compara con cada diferencia empírica de medias simple que tenga el diseño.

Si el valor de la diferencia de medias supera ese valor de rango crítico entonces se rechaza la hipótesis nula. Es decir, la diferencia entre el par de medias es estadísticamente significativa ($p < \alpha$) si dicha diferencia iguala o supera al valor de rango crítico que ofrece la prueba DHS de Tukey.

Los diferentes procedimientos de contrastes de hipótesis específicas se basan en la estimación de un valor de Rango Crítico obtenido a partir de una distribución concreta. En concreto, la prueba DHS de Tukey está basada en la distribución del rango estudentizado $q_{(\alpha, a, g|Error)}$, donde q es el valor de la distribución de rango estudentizado basado en el alfa, el número de medias a contrastar y los grados de libertad del error.

En el ejemplo, el valor del Rango Crítico es igual a 3.543. Por lo tanto, cualquier diferencia de medias que supere o iguale ese valor teórico (en valores absolutos) es estadísticamente significativa según la prueba de Tukey. En el

ejemplo las tres diferencias de medias superan el Rango Crítico de Tukey y, por lo tanto, se rechaza la hipótesis nula en los tres contrastes estadísticos de diferencias de medias.

Procedimiento de Dunnett

El procedimiento de Dunnett es el más potente cuando se trata de comparar:

- 1) la media de un grupo frente al resto de medias que tenga el diseño, es decir, se realizan $a-1$ comparaciones y además
- 2) son comparaciones simples.

No es un procedimiento exhaustivo sino que plantea hipótesis planificadas o a priori. También se calcula el rango crítico (RC) de Dunnett y se procede del mismo modo comparándolo con las diferencias de medias empíricas de la investigación.

La prueba de Dunnett utiliza la distribución basada en la comparación con un grupo de control (distribución de Dunnett). En la distribución de Dunnett se obtiene el valor de D en función del alfa, el número de grupos y los grados de libertad del error $D_{(\alpha, a, glError)}$. De nuevo, cualquier diferencia de medias (en valores absolutos) que iguale o supere dicho valor crítico es estadísticamente significativa.

Corrección de Bonferroni

La corrección de Bonferroni es el procedimiento más potente siempre que la hipótesis formule el número de comparaciones a priori aunque si C (número de comparaciones) es grande entonces la prueba es poco potente y no se recomienda su uso.

Se puede aplicar con hipótesis simples y con hipótesis complejas.

El procedimiento consiste en aplicar en cada comparación o prueba estadística el siguiente valor de alfa:

$Alfa = \alpha_{PE}$ que se desea en el experimento / Número de comparaciones (C)

$$\alpha_{PC} = \frac{\alpha_{PE}}{C}$$

Por ejemplo si se formulan 4 comparaciones o contrastes a priori, el α_{PE} final se mantendrá en .05 si en cada comparación individual se utiliza un alfa por comparación igual a .0125 (.05/4=.0125):

Si se ejecuta el procedimiento de Bonferroni se puede comprobar que $\alpha_{PE} = 1 - (1 - 0.0125)^4 = .049$. Valor cercano al alfa de .05 fijado a priori por el investigador.

El procedimiento de Bonferroni se basa en la distribución de la prueba $F: F_{(\alpha/C, 1, glError)}$, donde C representa el número de comparaciones y α/C es el alfa que se va a utilizar en cada comparación y por lo tanto la tabla que hay que consultar.

Procedimiento de Scheffé

El procedimiento de Scheffé es válido en cualquier circunstancia de investigación, pero normalmente es la prueba menos potente.

Se calcula el rango crítico (RC) de Scheffé y se procede del mismo modo comparándolo con las diferencias de medias empíricas. Si el valor de la diferencia de medias iguala o supera al valor del Rango Crítico entonces se rechaza la hipótesis nula.

La prueba de Scheffé está basada en la distribución de la prueba $F: F_{(\alpha, a-1, glError)}$.

En la figura 31 se detalla 'el máximo' número de contrastes (comparaciones) que deberían probarse en una investigación con el procedimiento de Bonferroni. Si el número de comparaciones es mayor entonces el procedimiento de Bonferroni pierde potencia estadística y sería más conveniente optar por la prueba de Scheffé que en esas circunstancias es más potente que Bonferroni.

Máximn nº de contrastes que deberían probarse con el procedimiento de Bonferroni

| glerror | Número de grupos | | | | | | | | |
|---------|------------------|---|----|----|-----|-----|-----|-----|--|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 5 | 2 | 4 | 8 | 12 | 17 | 24 | 31 | 40 | |
| 6 | 2 | 5 | 9 | 14 | 21 | 30 | 41 | 55 | |
| 7 | 2 | 5 | 10 | 16 | 25 | 37 | 52 | 71 | |
| 8 | 2 | 6 | 11 | 18 | 29 | 44 | 64 | 89 | |
| 9 | 2 | 6 | 12 | 20 | 33 | 51 | 75 | 107 | |
| 10 | 2 | 6 | 12 | 22 | 37 | 58 | 87 | 127 | |
| 12 | 3 | 7 | 13 | 25 | 43 | 70 | 110 | 166 | |
| 14 | 3 | 7 | 14 | 28 | 49 | 82 | 132 | 205 | |
| 16 | 3 | 7 | 15 | 30 | 54 | 93 | 153 | 243 | |
| 18 | 3 | 7 | 16 | 32 | 58 | 103 | 173 | 281 | |
| 20 | 3 | 7 | 17 | 33 | 63 | 112 | 191 | 316 | |
| 30 | 3 | 8 | 18 | 39 | 78 | 147 | 267 | 470 | |
| 40 | 3 | 8 | 20 | 43 | 87 | 170 | 320 | 586 | |
| 50 | 3 | 8 | 20 | 45 | 94 | 187 | 360 | 674 | |
| 60 | 3 | 8 | 21 | 47 | 98 | 199 | 390 | 743 | |
| 70 | 3 | 9 | 21 | 48 | 102 | 209 | 414 | 799 | |
| 80 | 3 | 9 | 21 | 49 | 105 | 217 | 433 | 844 | |
| 90 | 3 | 9 | 22 | 50 | 107 | 223 | 449 | 882 | |
| 100 | 3 | 9 | 2 | 50 | 109 | 228 | 462 | 913 | |
| 110 | 3 | 9 | 2 | 51 | 111 | 232 | 473 | 941 | |
| 120 | 3 | 9 | 22 | 51 | 112 | 236 | 483 | 964 | |

Figura 31. Máximo' número de contrastes con el procedimiento de Bonferroni

Siguiendo con el ejemplo anterior se pueden calcular las diferencias entre los pares de medias simples (ver los datos presentados anteriormente en la figura 28).

¿Qué prueba de contraste de hipótesis específicas será la más adecuada para contrastar todas las diferencias entre los pares de medias? La más adecuada es aquella que controla el error de Tipo I en el nivel fijado previamente por el alfa y además es la prueba con mayor potencia estadística (menor error de Tipo II). Por lo tanto, si se trata de un diseño donde se plantean todas las comparaciones de medias dos a dos (comparaciones exhaustivas y simples) entonces la prueba más potente es DHS de Tukey. La prueba DHS (diferencia honestamente significativa) de Tukey es la elección más correcta desde el punto de vista de la validez de conclusión estadística.

A continuación se detalla la salida de resultados que ofrece el programa SPSS (figura 32).

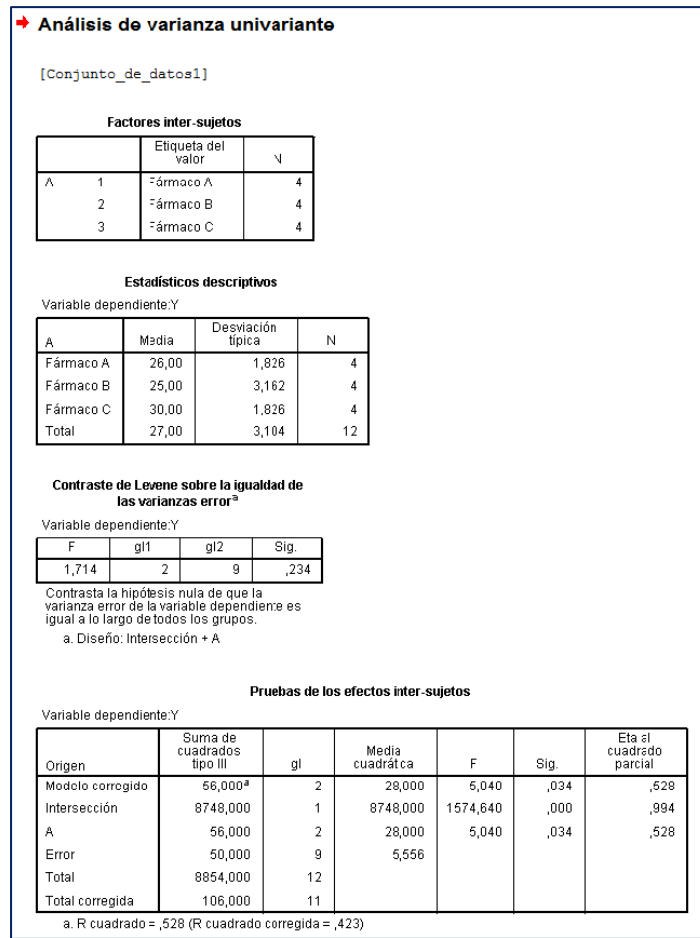


Figura 32. SPSS: resultados del ANOVA

Si se ejecuta el contraste de hipótesis mediante la prueba DHS de Tukey con el SPSS se seleccionará la opción de Tukey (figura 33).

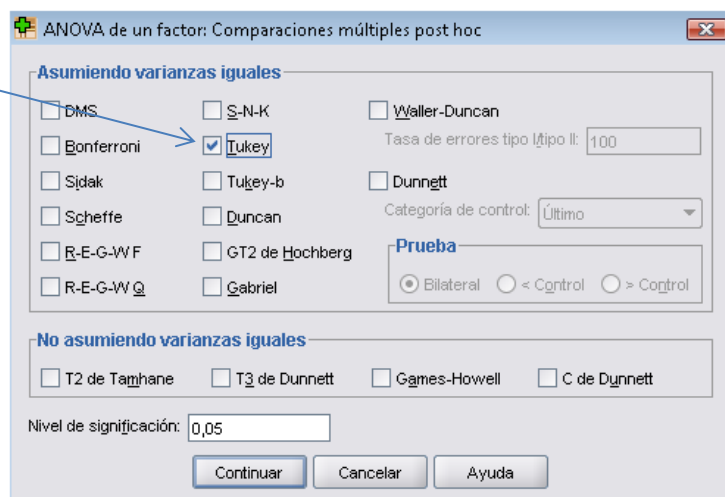


Figura 33. SPSS: resultados del ANOVA

Y el SPSS ofrece los siguientes resultados para la 'pruebas post hoc' HSD de Tukey (figura 34):

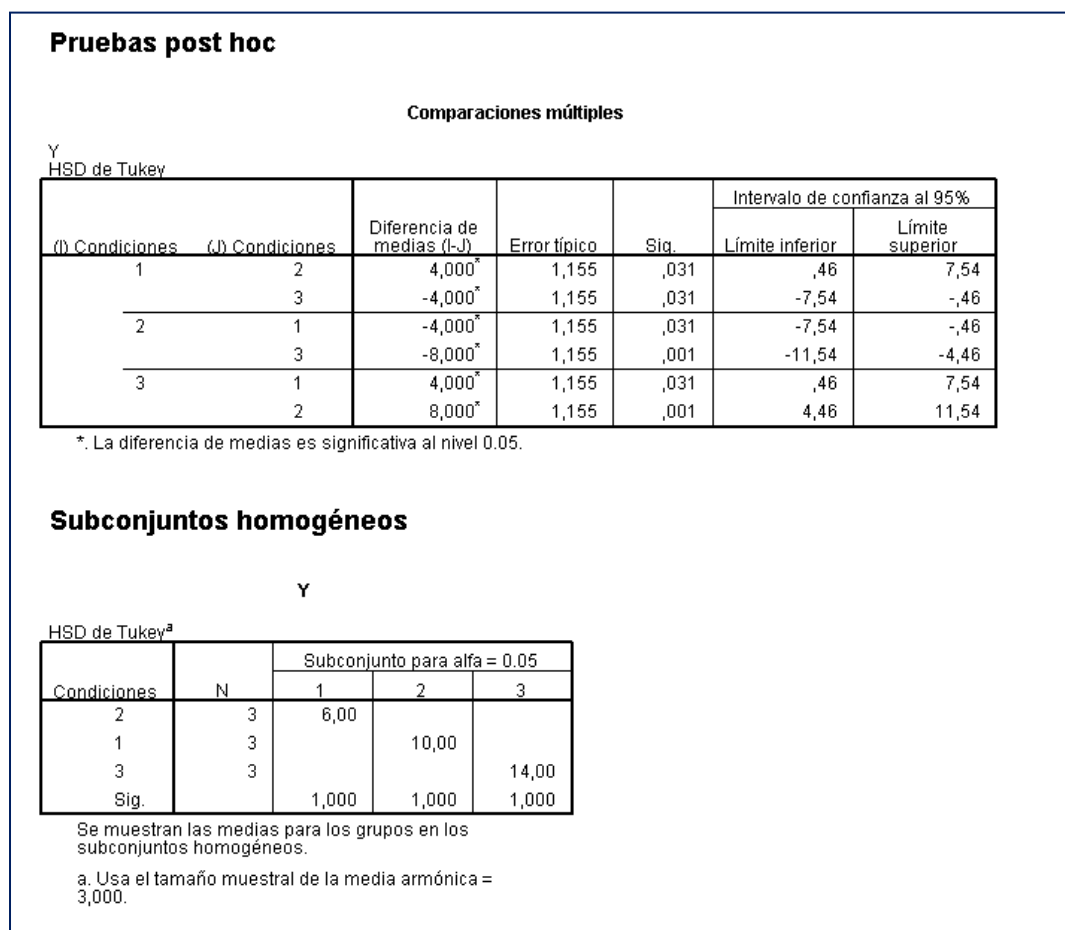


Figura 34. SPSS: resultados de la prueba de Tukey

La interpretación de los subconjuntos homogéneos que ofrece el SPSS es muy útil para visualizar de forma rápida si hay diferencias estadísticamente significativas entre las medias. Aquellas condiciones cuyas medias no difieren de forma estadísticamente significativa aparecen en la misma columna dentro del mismo subconjunto. Por el contrario, cuando dos medias difieren de forma estadísticamente significativa entre sí entonces aparecen en dos subconjuntos diferentes. En la tabla anterior se observa que cada media se encuentra en un subconjunto diferente indicando que las diferencias entre todos los pares de medias son estadísticamente significativas. Hay que revisar siempre la tabla de diferencias de medias, pues la prueba de subconjuntos homogéneos es más conservadora y podría ocurrir que no se detecten diferencias con el análisis visual de los subconjuntos y sí en la tabla de diferencias de medias de Tukey.

Si se hubiese optado por ejecutar la prueba de Bonferroni o Scheffé los resultados serían los siguientes (figura 35).

Pruebas post hoc

Comparaciones múltiples

Variable dependiente: Y

| (I) Condiciones | (J) Condiciones | Diferencia de medias (I-J) | Error típico | Sig. | Intervalo de confianza al 95% | | |
|-----------------|-----------------|----------------------------|--------------|-------|-------------------------------|-----------------|-------|
| | | | | | Límite inferior | Límite superior | |
| Scheffé | 1 | 2 | 4,000* | 1,155 | ,037 | ,30 | 7,70 |
| | | 3 | -4,000* | 1,155 | ,037 | -7,70 | -3,30 |
| | 2 | 1 | -4,000* | 1,155 | ,037 | -7,70 | -3,30 |
| | | 3 | -8,000* | 1,155 | ,001 | -11,70 | -4,30 |
| | 3 | 1 | 4,000* | 1,155 | ,037 | ,30 | 7,70 |
| | | 2 | 8,000* | 1,155 | ,001 | 4,30 | 11,70 |
| Bonferroni | 1 | 2 | 4,000* | 1,155 | ,040 | ,20 | 7,80 |
| | | 3 | -4,000* | 1,155 | ,040 | -7,80 | -2,20 |
| | 2 | 1 | -4,000* | 1,155 | ,040 | -7,80 | -2,20 |
| | | 3 | -8,000* | 1,155 | ,001 | -11,80 | -4,20 |
| | 3 | 1 | 4,000* | 1,155 | ,040 | ,20 | 7,80 |
| | | 2 | 8,000* | 1,155 | ,001 | 4,20 | 11,80 |

*. La diferencia de medias es significativa al nivel 0.05.

Figura 35. SPSS: resultados de la prueba de Bonferroni y Scheffé

La prueba de Scheffé es más potente (menor error de Tipo II) que Bonferroni ya que si se observa la figura anterior de ‘máximo número de contrastes (comparaciones) que deberían probarse con el procedimiento de Bonferroni’ (ver la figura anterior con la tabla en la figura 31) se puede comprobar que Bonferroni sería más potente si el máximo de comparaciones fuese de dos contrastes para un diseño con tres grupos y seis grados de libertad del error.

De todos modos para esa situación donde se realizan todas las comparaciones posibles dos a dos ya se ha comprobado que la prueba más potente es DHS de Tukey (valores de $p = .031$ y $p = .001$). Y esa sería la decisión más correcta para ese diseño y, por lo tanto, la que debería seleccionar el investigador o investigadora para realizar sus hipótesis específicas, asegurando la validez de conclusión estadística.

En aquellas situaciones de investigación donde la prueba más adecuada es la de Dunnett (donde se realizan $a-1$ comparaciones simples) entonces será necesario indicar en el SPSS qué grupo es el de comparación (categoría de control) para que efectúe las $a-1$ comparaciones respecto a un determinado grupo. Por ejemplo, si la categoría de control se llama ‘Primero’ entonces eso significa que se

va a comparar la media del primer grupo con las medias del resto de condiciones que tenga el diseño (figura 36). En un diseño con $A = 3$ las comparaciones que efectúa la prueba de Dunnett serían: $a_1 - a_2$ y $a_1 - a_3$.

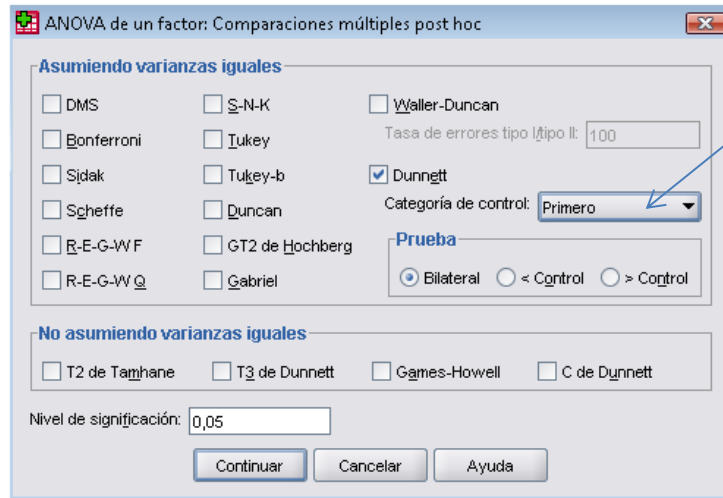


Figura 36. SPSS: selección de la prueba de Dunnett

Suponiendo que el investigador o investigadora desea realizar $a - 1$ comparaciones simples ($3 - 1 = 2$ contrastes) entonces la prueba de Dunnett ofrece el siguiente resultado en el SPSS (figura 36). Como se observa, sólo aparecen dos contrastes: la media del grupo a_2 con la media del grupo a_1 (la diferencia de medias es -4) y la media del grupo a_3 con la media del grupo a_1 (diferencia de medias de 4). Los resultados que ofrece el SPSS para poder tomar la decisión estadística de si se puede o no rechazar la hipótesis nula es el valor de significación (Sig.), es decir, el valor p de probabilidad del estadístico de comparación.

Pruebas post hoc

Comparaciones múltiples

Y
t de Dunnett (bilateral)^a

| (I) Condiciones | (J) Condiciones | Diferencia de medias (I-J) | Error típico | Sig. | Intervalo de confianza al 95% | |
|-----------------|-----------------|----------------------------|--------------|------|-------------------------------|-----------------|
| | | | | | Límite inferior | Límite superior |
| 2 | 1 | -4,000* | 1,155 | ,024 | -7,31 | -,69 |
| 3 | 1 | 4,000* | 1,155 | ,024 | ,69 | 7,31 |

a. Las pruebas t de Dunnett tratan un grupo como control y lo comparan con todos los demás grupos.
*. La diferencia de medias es significativa al nivel 0.05.

Figura 37. SPSS: resultados de la prueba de Dunnett

En ningún tipo de prueba de hipótesis específicas informa el SPSS del valor del Rango Crítico del estadístico aplicado sino que directamente informa del valor p de probabilidad vinculado al contraste de diferencia de medias. Otros programas estadísticos sí informan del valor p de probabilidad junto al valor de Rango Crítico como por ejemplo la aplicación del ANOVA que se encuentra en la Web Vassar Stats (<http://vassarstats.net/>).

La ejecución de un modelo de diseño entre-sujetos unifactorial $A = 3$ univariado mediante la Web Vassar Stats se detalla a continuación (figura 37). Se puede observar que sus resultados sí aportan el valor del Rango Crítico de la prueba de Tukey. El lector o lectora puede comprobar el valor del rango crítico para dos valores de alfa: $RC_{.05} = 4.66$ y $RC_{.01} = 6.4$. A medida que disminuye el valor p de probabilidad, aumenta el valor del Rango Crítico, tal y como sucede con los valores de la F teórica o de cualquier estadístico de contraste. Una vez se conoce el valor del Rango Crítico se puede comparar con la diferencia entre las medias. Si la diferencia de medias es \geq que el valor del Rango Crítico entonces se concluye que la diferencia hallada es estadísticamente significativa. Si la diferencia de medias es $<$ que el valor del Rango Crítico entonces se concluye que la diferencia hallada no es estadísticamente significativa.

ANOVA UNIFACTORIAL A=3 CON VASSAR STATS

<http://vassarstats.net/>

Cuando el número de muestras es $k=2$, el análisis de varianza (análisis estándar con ponderación de medias) es equivalente a una Prueba t con $F=t^2$.

Primer paso

número de muestras en el análisis = 3

Seleccione esta opción solo si desea realizar un análisis no-ponderado. Aviso: Seleccione esta opción sólo si tiene razones fundadas para ello.

Seleccione para realizar un análisis ponderado

Ingreso de datos

| Muestra 1 | Muestra 2 | Muestra 3 | Muestra 4 | Muestra 5 |
|-----------|-----------|-----------|-----------|-----------|
| 28 | 28 | 29 | | |
| 24 | 27 | 31 | | |
| 27 | 21 | 28 | | |
| 25 | 24 | 32 | | |

| Resumen de datos | | | | | | |
|---------------------|----------|--------|--------|---|---|--------|
| | Muestras | | | | | Total |
| | 1 | 2 | 3 | 4 | 5 | |
| N | 4 | 4 | 4 | | | 12 |
| ΣX | 104 | 100 | 120 | | | 324 |
| Mean | 26 | 25 | 30 | | | 27 |
| ΣX^2 | 2714 | 2530 | 3610 | | | 8854 |
| Varianza | 3.3333 | 10 | 3.3333 | | | 9.6364 |
| Desviación estándar | 1.8257 | 3.1623 | 1.8257 | | | 3.1042 |
| Error Estándar | 0.9129 | 1.5811 | 0.9129 | | | 0.8961 |

| análisis ponderado de promedios estándar | | | | | |
|--|------------|--------------------|--------|------|-----------------------------|
| Resumen ANOVA muestras independientes I | | | | | |
| Fuente | SS | Grados de Libertad | MS | F | P |
| Tratamiento [entre grupos] | 56 | 2 | 28 | 5.04 | 0.034001 |
| Error | 50 | 9 | 5.5556 | | |
| Ss/BI | | | | | Graph Maker |
| Total | 106 | 11 | | | |

Ss/BI = Sujetos o bloques según el diseño.
Solamente aplicable a ANOVA para muestras relacionadas.

Prueba HSD Tukey
 HSD[.05]=4.66; HSD[.01]=6.4
 M1 vs M2 no-significativo
 M1 vs M3 no-significativo
 M2 vs M3 P<.05

M1 = promedio de muestra 1
 M2 = promedio de muestra 2
 and so forth.

HSD = diferencia absoluta entre los promedios de cualquiera de las muestras requerida para obtener una diferencia significativa en algún nivel. HSD[.05] para un nivel de .05. HSD[.01] para un nivel de .01.

Figura 38. Aplicación on line de Vassar Stats: diseño entre-grupos A = 3

Análisis con el programa SPSS

En la figura 38 se reproducen los resultados de un estudio tal y como se obtienen con el programa SPSS junto con la ventana de datos que se ha elaborado para introducirlos en la base cuando se trata de un *diseño entre-grupos unifactorial univariado* (en la ventana de datos poner las etiquetas al factor A: situar 1 para shock escapable y 2 para shock inescapable).

Para llevar a cabo el análisis con el SPSS se seleccionan las siguientes instrucciones una vez nos encontramos siguiendo el siguiente recorrido en las ventanas del programa:

ANALIZAR → Modelo lineal general → univariado

A continuación se sitúan en la base las variables en su lugar correcto:

-variable dependiente Y

-factores fijos: variable independiente A

Y también se selecciona, a la derecha, en la ventana de 'Opciones' los apartados de: estadísticos descriptivos, estimaciones del tamaño del efecto y pruebas de homogeneidad.

Una vez se ejecutan esas instrucciones, los resultados que aporta el programa relativos al análisis de la varianza solicitado se detallan en la figura 38. Esos resultados del ANOVA y sus estadísticos deben ser redactados en el informe de investigación tal y como se ha detallado anteriormente.

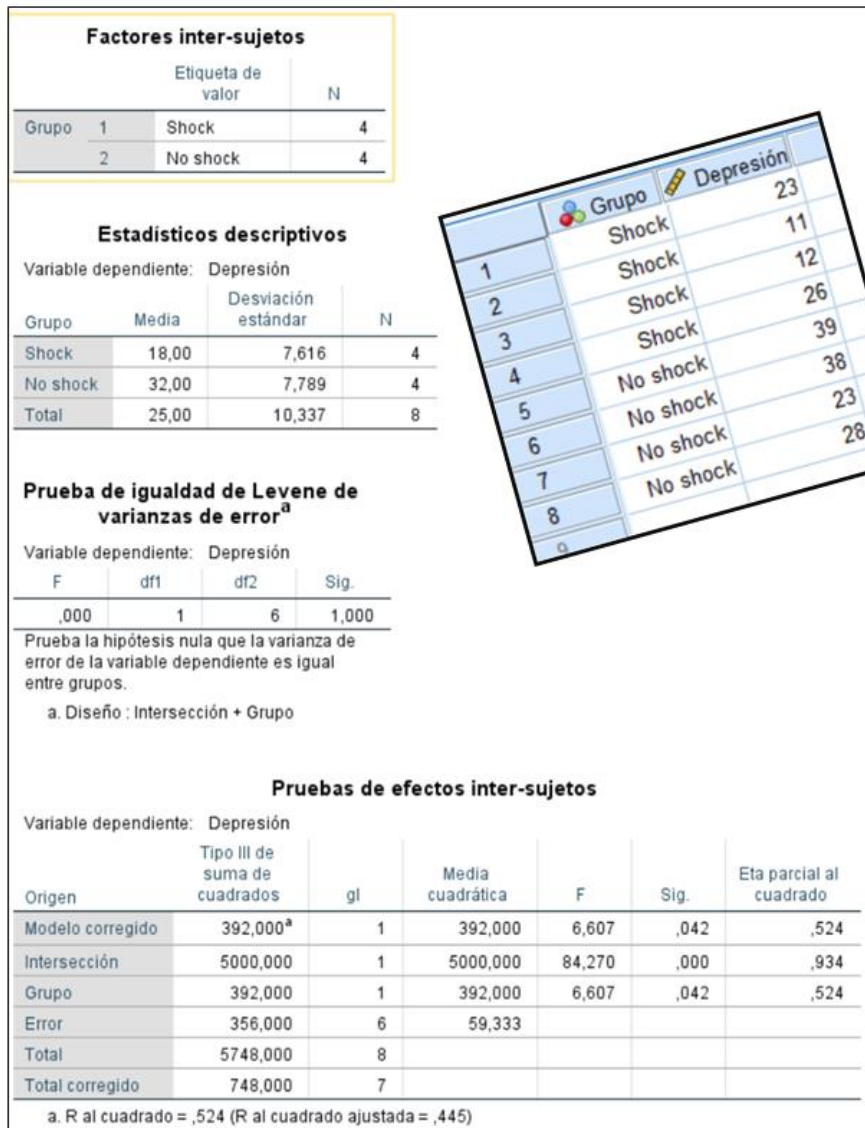


Figura 39. Datos introducidos en el SPSS y resultados del ANOVA

SPSS: ANOVA de un factor para muestras o grupos independientes

A continuación se presenta un conjunto de datos para un diseño entre grupos unifactorial univariado $A=2$ analizado con la prueba de contraste de hipótesis de ANOVA (Análisis de la Varianza) mediante el programa estadístico SPSS. Los datos corresponden a un diseño entre-sujetos o entre-grupos con una única variable independiente o factor (A). El diseño 'entre-sujetos' señala que cada sujeto solamente es medido en una ocasión ya sea expuesto a la condición a_1 del factor o variable independiente o a la condición a_2 de dicho factor. También se conoce como un 'diseño de muestras independientes'. Además es un diseño con una sola variable dependiente o variable medida (Y) y por ello es un diseño 'univariado'. La introducción de los datos de un diseño entre-grupos unifactorial univariado en el programa de SPSS se realiza tal y como se detalla en la figura 39.

| | Condición | Y | var |
|---|-----------|----|-----|
| 1 | 1 | 13 | |
| 2 | 1 | 5 | |
| 3 | 1 | 11 | |
| 4 | 1 | 11 | |
| 5 | 2 | 27 | |
| 6 | 2 | 21 | |
| 7 | 2 | 32 | |
| 8 | 2 | 24 | |
| 9 | | | |

Figura 40. Datos introducidos en el SPSS y resultados del ANOVA

A continuación se puede ejecutar el análisis de la prueba estadística de ANOVA para un factor (figura 40). En el SPSS se selecciona la ventana de:

ANALIZAR → comparar medias → ANOVA de un factor

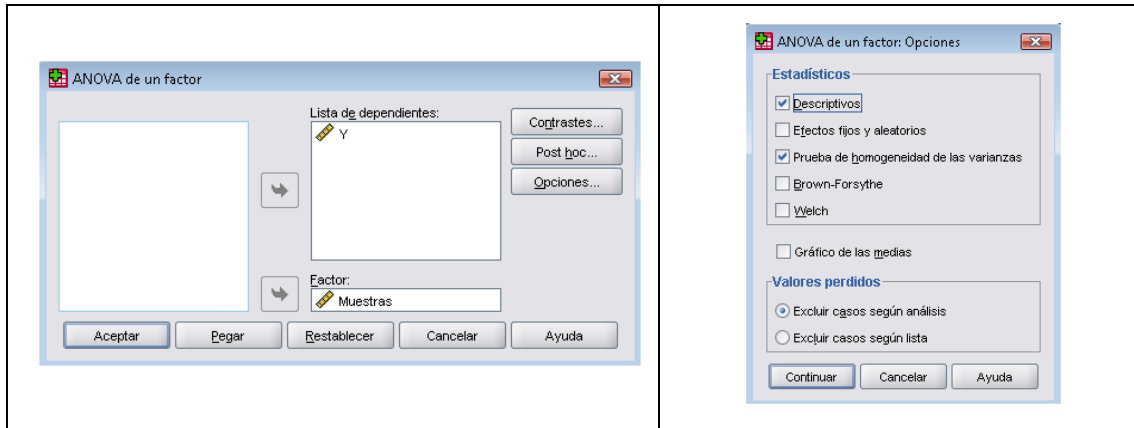


Figura 41. Datos introducidos en el SPSS y resultados del ANOVA

Dentro de 'ANOVA de un factor' seleccionamos Opciones y clicamos sobre Estadísticos: Descriptivos y Prueba de homogeneidad de las varianzas de los dos grupos. Se aconseja realizar el análisis con la ventana de Modelo Lineal General ya que en ese apartado se encuentran todos los modelos de diseño.

Los resultados del ANOVA de un diseño entre-grupos A= 2 univariado son los siguientes (figura 41):

ANOVA de un factor

[Conjunto_de_datos1]

Descriptivos

Y

| | N | Media | Desviación típica | Error típico | Intervalo de confianza para la media al 95% | | Mínimo | Máximo |
|-------|---|-------|-------------------|--------------|---|-----------------|--------|--------|
| | | | | | Límite inferior | Límite superior | | |
| 1 | 4 | 10,00 | 3,464 | 1,732 | 4,49 | 15,51 | 5 | 13 |
| 2 | 4 | 26,00 | 4,690 | 2,345 | 18,54 | 33,46 | 21 | 32 |
| Total | 8 | 18,00 | 9,366 | 3,311 | 10,17 | 25,83 | 5 | 32 |

Prueba de homogeneidad de varianzas

Y

| Estadístico de Levene | gl1 | gl2 | Sig. |
|-----------------------|-----|-----|------|
| ,429 | 1 | 6 | ,537 |

ANOVA

Y

| | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
|--------------|-------------------|----|------------------|--------|------|
| Inter-grupos | 512,000 | 1 | 512,000 | 30,118 | ,002 |
| Intra-grupos | 102,000 | 6 | 17,000 | | |
| Total | 614,000 | 7 | | | |

Figura 42. Datos introducidos en el SPSS y resultados del ANOVA

Los resultados señalan, por un parte, que sí existe homogeneidad de las varianzas ($Levene_{(1, 6)} = 0.429, p = .537$). Y, por otra parte, la prueba de contraste estadístico del Análisis de la Varianza (ANOVA) señala que la diferencia entre las medias del grupo 1 (Media = 10, DT=3.46) y del grupo 2 (Media = 26, DT=4.69) es estadísticamente significativa ($F_{(1, 6)} = 30, 118, p = .002$). Por lo tanto, la media del grupo 2 es mayor que la media del grupo 1, siendo la diferencia entre las medias estadísticamente significativa.

Se puede establecer una relación directa entre el valor del estadístico de contraste t de Student y el valor del estadístico de contraste F del Análisis de la Varianza dado que $t = F^2$ y $F = \sqrt{t}$.

Análisis con el programa JASP

A continuación, en la figura 42 se detalla la base de datos que se utiliza con el programa gratuito JASP y el resultado que se obtiene. La introducción de datos no se realiza en el programa JASP y por ello se necesitan otros programas. La introducción de los datos se puede realizar con Excel o con SPSS y luego se llama a ese archivo de datos desde el programa JASP. Para ello, se abre el programa JASP y se selecciona en la izquierda donde están situadas tres líneas verticales y se busca en el ordenador al fichero que se desea abrir:

Open → Computer

Así, se busca en el ordenador el fichero de datos que se va a utilizar para el contraste de hipótesis estadísticas y se selecciona para que se abra en el programa JASP.

Cuando ya se dispone en la pantalla del ordenador los datos que son objeto de análisis, se selecciona la ventana que tiene la opción de ANOVA y se van seleccionando las casillas que se corresponden con el diseño que se va a ejecutar: *diseño entre-grupos unifactorial univariado*. Es decir, al abrir la ventana se selecciona 'ANOVA' y se sitúan las variables dependiente ('Depresión') e independiente (factor fijo) ('Grupo') (figura 42).

Además, se seleccionan las opciones de:

-1) 'Assumption Checks: Homogeneity tests' para analizar si las varianzas de los dos grupos son homogéneas.

2) Marginal Means: Descriptive statistics, Estimates of effect size y η^2 .

El estadístico η^2 es el tamaño del efecto conocido como eta cuadrado o proporción de la varianza explicada por el efecto del tratamiento o Grupo. Así, se calcula como:

$$\eta^2 = \text{Suma de Cuadrados del efecto} / \text{Suma de Cuadrados Total.}$$

Su valor complementario, $1 - \eta^2$, es la proporción de varianza no explicada por dicho efecto y que se corresponde con la varianza explicada por el término de error en la ecuación estructural planteada en este diseño de investigación:

$$1 - \eta^2 = \text{Suma de Cuadrados del error} / \text{Suma de Cuadrados Total.}$$

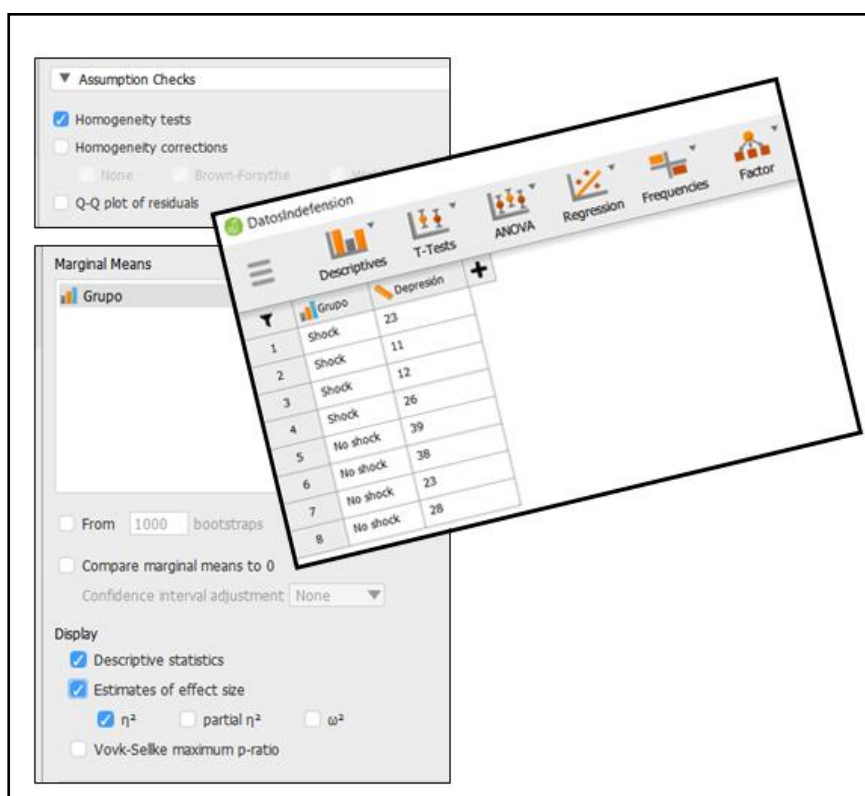


Figura 43. Datos introducidos en JASP y resultados del ANOVA

Los resultados de las instrucciones dadas al programa JASP se representan en la figura 43.

ANOVA

ANOVA - Depresión

| Cases | Sum of Squares | df | Mean Square | F | p | η^2 |
|----------|----------------|------|-------------|------|------|----------|
| Grupo | 392.00 | 1.00 | 392.00 | 6.61 | 0.04 | 0.52 |
| Residual | 356.00 | 6.00 | 59.33 | | | |

Note. Type III Sum of Squares

Assumption Checks

Test for Equality of Variances (Levene's)

| F | df1 | df2 | p |
|-----------|------|------|------|
| 7.95e -30 | 1.00 | 6.00 | 1.00 |

Descriptives

Descriptives - Depresión

| Grupo | Mean | SD | N |
|----------|-------|------|---|
| Shock | 18.00 | 7.62 | 4 |
| No shock | 32.00 | 7.79 | 4 |

Figura 44. Resultados del ANOVA realizado con JASP

Una de las ventajas que tiene el programa JASP (además de ser un programa gratuito que se descarga de su página web (<https://jasp-stats.org/>) y que constantemente se está actualizando porque va incorporando más posibilidades de análisis) es que permite estimar el denominado Factor Bayes, FB. El Factor Bayes (BF) estima la probabilidad de una hipótesis estadística sobre la probabilidad de la otra hipótesis como ya se comentó anteriormente. En este caso, el usuario selecciona:

ANOVA → Bayesian ANOVA

Una vez seleccionado al ANOVA bayesiano, se sitúa la variable dependiente y la variable independiente en su lugar correcto y se marca la opción de BF10 si se desea estimar la probabilidad de la hipótesis alternativa respecto a la probabilidad de la hipótesis nula (figura 44). Se podría marcar la opción de BF01 y, entonces, el programa ofrece el resultado de la probabilidad de la hipótesis nula respecto a la probabilidad de la hipótesis alternativa.

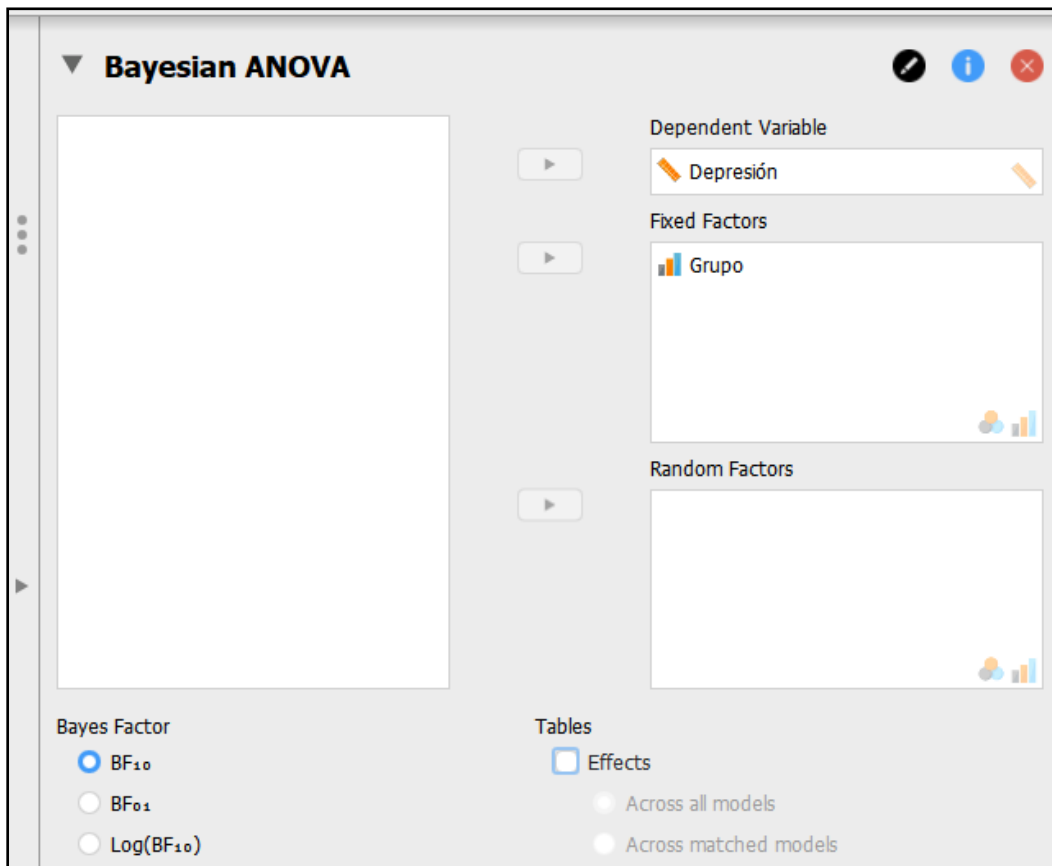


Figura 45. Factor Bayes con JASP

A continuación el programa JASP ofrece los resultados que se detallan en la figura 45. El valor del Factor Bayes es: $BF_{10} = 2.11$, es decir, la hipótesis alternativa es 2.11 más probable que la hipótesis nula.

Results

Bayesian ANOVA

Model Comparison

| Models | P(M) | P(M data) | BF_M | BF_{10} | error % |
|------------|------|-----------|--------|-----------|----------|
| Null model | 0.50 | 0.32 | 0.47 | 1.00 | |
| Grupo | 0.50 | 0.68 | 2.11 | 2.11 | 1.04e -4 |

Figura 46. Factor Bayes con JASP

Referencias

American Psychological Association (1953). *Ethical standards of psychologists*. Washington, DC: American Psychological Association.

American Psychological Association (2001). *Publication Manual of the American Psychological Association* (5th Ed.). Washington, DC: American Psychological Association.

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060-1073.

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (6th Ed.). Washington, DC: American Psychological Association.

American Psychological Association (2019). *Publication Manual of the American Psychological Association* (7th Ed.). Washington, DC: American Psychological Association.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in Psychology Why do we need them? What might they be? *American Psychologist*, 63, 839-851.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal Article Reporting Standards for quantitative research in Psychology: The APA Publications and Communications Board Task Force report. *American Psychologist*, 73, 3-25.

Arnau, J. (1981). *Diseños experimentales en psicología y educación*. Vol. I. México D.F.: Trillas (Preedición).

Ato, M. (1991). *Investigación en Ciencias del Comportamiento*. I: Fundamentos. Barcelona: PPU.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researcher misunderstand confidence intervals and standard error bars. *Psychological Methods*, 10, 389-396.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building* (1st ed.). New York, NY: Wiley & Sons.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (2005). *Statistics for experimenters: design, innovation and discovering* (2nd ed.). New York, NY:Wiley & Sons.

Código Deontológico del Psicólogo (2010). Colegio Oficial de Psicólogos de Madrid. Disponible en:

<http://www.copmadrid.org/webcopm/recursos/codigodeontologicojunio2010.pdf>

Cumming, G. & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61, 532-575.

Cumming, G. & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60, 170-180.

delMas, R.C. & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4, 55-82.

Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning*, 9, 83-96.

Frías-Navarro, D. & Pascual-Llobell, J. (2003). Psicología clínica basada en pruebas: efecto del tratamiento. *Papeles del Psicólogo*, 85, 11-18.

Frías-Navarro, D. (2011). *Técnica estadística y diseño de investigación*. Valencia: Palmero Ediciones.

Frías-Navarro, D., Pascual-Llobell, J., & García, J.F. (2000). La hipótesis nula y la significación práctica. *Revista de la Asociación Española de Metodología de las Ciencias del Comportamiento*, Volumen especial, 181-185.

Frías-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perez-Gonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production? *European Journal of Education*. First published: 18 August 2020. Doi: <https://doi.org/10.1111/ejed.12417>

García de la Banda, G., Martínez-Abascal, M.A., Riesco, M., & Pérez G. (2004). La respuesta de cortisol ante un examen y su relación con otros acontecimientos estresantes y con algunas características de personalidad. *Psicothema*, 16, 294-298.

García, J.F., Frías-Navarro, D., & Pascual, J. (2006). *Los diseños de la investigación experimental. Comprobación de las hipótesis*. Valencia: editorial CSV. (Incluye CD con ejercicios).

Gisberta, J. P. & Bonfill, X (2004). ¿Cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis? *Gastroenterología y Hepatología*, 27, 129-149.

González de Dios, J., González-Muñoz, M., Alonso-Arroyo, A., y Benavent, A. (2014). Comunicación científica (XVIII). Conocimientos básicos para leer (y escribir) un artículo científico (5): listas de comprobación de documentos. *Acta Pediátrica Española*, 72(11), e389-e392.

Hancock, G.R. & Mueller, R.O. (2010). *The Reviewer's guide to quantitative methods in the Social Sciences*. New York, NY: Taylor & Francis.

Hoekstra, R., Finch, S., Kiers H. A. L., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of p values. *Psychonomic Bulletin & Review*, 13, 1033-1037.

Huck, S. W. (2007). Reform in statistical education. *Psychology in the School*, 44, 527-533.

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings among experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A*, 171, 481-502.

Ioannidis, J. P. A. (2018) Meta-research: Why research on research matters. *PLoS Biology*, 16(3): e2005468, 1-6.

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLoS Biology*, 13(10): e1002264, 1-7.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York, NY: Holt, Rinehart and Winston.

Kirk, R.E. (1995). *Experimental design: procedures for the behavioral sciences*. (3rd Ed.). Belmont, CA: Brooks/Cole Publishing.

Kish, L. (1975). *Diseño estadístico para la investigación* (versión en castellano, 1995). Madrid: Centro de Investigaciones Sociológicas (CIS), Colección Monografías N° 146.

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal Article Reporting Standards for qualitative primary, qualitative meta-analytic, and mixed methods research in Psychology: The APA Publications and Communications Board Task Force report. *American Psychologist*, 73, 26-46.

Maxwell, S. E. & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd Ed.). Mahwah, NJ: Erlbaum.

Maxwell, S. E., Delaney, H.D., & Kelley (2018). *Designing experiments and analyzing data: A model comparison perspective* (3rd ed.). New York, NY: Routledge

Mayor, J. & Pérez, J. (1989). Psicología o psicologías? Un problema de identidad. En J. Mayor y J. L. Pinillos (Eds.), *Tratado de psicología general*, vol. 1: Historia, teoría y método. (pp. 3-69). Madrid: Alhambra Universidad.

Mood, A. M. & Graybill, F. A. (1972). *Introducción a la teoría de la estadística*. Madrid: Aguilar ediciones.

Nesselroade, J. R. & Cattell, R. B. (1988) (Eds.). *Handbook of multivariate experimental Psychology*. (2nd ed. rev.). New York, NY: Plenum Press.

Onwuegbuzie, A. J. (2001). Common methodological, analytical, and interpretational errors in published educational studies: An analysis of the 1998 volume of the British Journal of Educational Psychology. *Educational Research Quarterly*, 26, 11-22.

Pascual-Llobell, J., Frías-Navarro, D., & García-Pérez, J. F. (1996). *Manual de Psicología Experimental*. Barcelona: Ariel.

Pascual-Llobell, J., Frías-Navarro, D., & García-Pérez, J. F. (2000). El procedimiento de significación estadística (NHST): su trayectoria y actualidad. *Revista de Historia de la Psicología*, 21, 9-26.

Pascual-Llobell, J., Frías-Navarro, D., & Monterde-i-Bort, H. (2004). Tratamientos psicológicos con apoyo empírico y práctica basada en la evidencia. *Papeles del Psicólogo*, 87, 1-8.

Rosnow, R.L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276–1284.

Sackett, D. L., Strauss, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd Ed.). Edinburgh: Churchill Livingstone.

Sánchez-Meca, J. & Botella, J. (2010). Revisiones sistemáticas y meta-análisis: Herramientas para la práctica profesional. *Papeles del Psicólogo*, 31, 7-17.

Santiago, M. I., Hervada, X., Naveira, G., Silva, L. C., Fariñas, H., Bacallao, J., & Mújica, O. J. (2010). El programa epidat: usos y perspectivas. *Revista Panamericana de Salud Pública*, 27, 80-82.

Shaughnessy, J. J. (2009). Evaluating and understanding articles about treatment. *American Family Physician*, 79, 668-670.

Smaldino, P.E. & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3: 160384, 1-17.

Vacha-Haase, T., Nilsson, J.E., Reetz, D.R., Lance, T.S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology*, 10, 413-425.

Wilkinson, L., and the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.

ANEXOS

Anexo 1. Breve explicación de conceptos fundamentales de diseño de investigación

Diseño de entre-grupos o entre-sujetos (between subjects or between groups design).

Se trata de un diseño de investigación donde el investigador:

1) asigna un determinado grupo o condición que tiene la variable independiente manipulada o factor (por ejemplo, el tratamiento nuevo o el tratamiento tradicional) a los participantes del estudio. Si la asignación de la condición es aleatoria se trata de un estudio con una metodología experimental y si no es aleatoria se trata de una metodología cuasi-experimental, o bien

2) selecciona a los participantes de determinados grupos independientes o diferentes que ya están formados ('grupos intactos') en función de sus características o propiedades (por ejemplo, vivir en zona rural o urbana, ser hombre o ser mujer...). Se trata de un diseño entre-grupos realizado con una metodología no experimental.

Diseño de medidas repetidas o diseño intra-sujetos (repeated measures design or within-subjects design).

Se trata de un diseño de investigación que implica que todos los participantes son medidos en los diferentes niveles o condiciones del factor (también conocido como variable independiente).

En ocasiones, estos diseños suponen que los participantes se miden en diferentes momentos temporales (por ejemplo, pre-test y post-test e incluso seguimiento) o pueden suponer que los participantes reciben los diferentes tratamientos que se prueban en el estudio (por ejemplo, reciben el fármaco denominado Z y también el fármaco denominado W y sus efectos se miden en una determinada variable).

Diseño mixto o de medidas parcialmente repetidas (mixed design or design with both between and within-subjects factors).

Se trata de un diseño que al menos tiene un factor entre-grupos (factor A, por ejemplo) y un factor de medidas repetidas (factor B, por ejemplo).

El diseño mixto A x B tiene dos términos de error:

1) un error formado por los sujetos (S) anidados en una condición del factor entre-grupos A (Error = S / A) para calcular el denominador de la razón *F* de las fuentes de varianza denominadas *entre-grupos* (factor A entre-grupos). Sería el término de error que se utiliza para calcular la razón *F* del factor A.

2) un error formado por la interacción entre el factor B de medidas repetidas y el factor S de Sujetos que a su vez está anidado en el factor entre-sujetos A (Error = B x S / A) para las fuentes de varianza de medidas repetidas y para las interacciones que la variable de medidas repetidas tenga con los factores entre-grupos.). Sería el término de error que se utiliza para calcular la razón *F* del factor B y de la interacción A x B.

Diseño de bloques (block design).

Se trata de un diseño con al menos una variable bloqueada (variable categorizada en condiciones). Cuando se estudia la relación entre una variable independiente y la variable dependiente es frecuente que aparezcan otras variables o factores que también influyen y que deben ser controladas en el diseño de investigación. A estas variables se las denomina 'variables bloque' o 'variables de bloqueo', y se caracterizan por 1) no ser el motivo del estudio (es decir no están formuladas en la hipótesis de investigación, no son variables explicativas) sino que forman parte del estudio dado que tienen un efecto estadísticamente significativo sobre las puntuaciones de la variable dependiente y 2) porque se asume que no tienen interacción con el factor o variable independiente del estudio que forma parte de la hipótesis de investigación); es decir, se debe cumplir el supuesto de que el efecto de la interacción entre la variable independiente de la hipótesis y la variable de boqueo no es estadísticamente significativo.

El modelo es de "bloques aleatorizados completos" cuando en cada bloque se presentan todos los posibles tratamientos del factor objeto de estudio y, además,

dentro de cada bloque se asignan los tratamientos de forma aleatoria. Podría ser de bloques no aleatorizados si no hay asignación aleatoria del tratamiento a cada bloque. En ocasiones también puede ocurrir que no se pueden asignar todos los tratamientos sobre cada bloque, de modo que se tienen los diseños por bloques aleatorizados (o no aleatorizados) incompletos.

En el diseño completamente aleatorizado se asignan los tratamientos al azar a los grupos, sin restricción alguna, mientras que en el diseño en bloques aleatorizados primero se forman niveles o condiciones de los bloques (las denominadas parcelas) y a continuación se asignan los tratamientos (las condiciones) a las parcelas de cada bloque. Por lo tanto, se trata de un diseño con aleatorización restringida, ya que las unidades experimentales (sujetos) son primero clasificadas en grupos homogéneos (bloques), y después los tratamientos o condiciones son asignados aleatoriamente dentro de los bloques.

Diseño con variables covariadas, ANCOVA (Analysis of Covariance, ANCOVA).

Se trata de un diseño que al menos tiene una variable covariada (variable medida cuantitativa continua) que se introduce en el diseño como una fuente de varianza más, pero dicha variable no forma parte de la hipótesis de investigación. La variable covariada se puede utilizar para controlar una variable extraña, para reducir el error o para producir ambas cosas.

Se conoce como ANCOVA cuando el diseño solo tiene una variable dependiente y se trata de un MANCOVA cuando el diseño tiene más de una variable covariada. Se trata de un diseño análogo al diseño de bloques y la diferencia está en que en el ANCOVA los participantes se aparean matemáticamente en lugar de hacerlo físicamente formando bloques con participantes que tienen características similares en la variable de bloqueo (diseño de bloques) (Huitema, 1980; Reichardt, 1979).

En el ANCOVA se realiza un ajuste matemático de la variabilidad observada en los datos del estudio a través de las puntuaciones que se obtienen en la variable continua covariada.

Diseño de grupo equivalente o diseño con grupo de control equivalente (equivalente group design).

Se trata de un diseño elaborado con una metodología experimental. Este diseño permite comparar grupos de participantes o unidades experimentales que son asignados a los diferentes tratamientos o intervenciones de una manera aleatoria (por azar). De ahí que se denomine diseño de 'grupos equivalentes', ya que se utiliza la técnica de la aleatorización para mantener controladas por azar las posibles diferencias entre los grupos antes de introducir el tratamiento o las diferentes condiciones de intervención. Los sujetos de los grupos son homogéneos o equivalentes en las variables observadas y no observadas antes de introducir la variable manipulada o variable independiente. La principal técnica de control que se utiliza es la aleatorización en la asignación del tratamiento, pero si hay dudas de la calidad de esta medida se pueden aplicar también otras técnicas de control para controlar posibles variables extrañas conocidas por el investigador.

Diseño de grupo no equivalente o diseño con grupo de control no equivalent (nonequivalente group design).

Se trata de un diseño con una metodología cuasi-experimental. Este diseño permite comparar grupos de participantes o unidades experimentales que son asignados a los diferentes tratamientos o intervenciones de una manera no aleatoria. De ahí que se denomine diseño de 'grupos no equivalentes'. Los sujetos de los grupos no son homogéneos o equivalentes en las variables observadas y no observadas antes de introducir la variable manipulada o variable independiente y, por lo tanto, requiere planificar técnicas de control dirigidas a controlar posibles variables extrañas que el investigador considere relevantes. Por ejemplo, la eliminación, el bloqueo o la medida de una variable covariada.

Análisis de la Varianza (ANOVA, Analysis of Variance).

ir Ronal Fisher utilizó por primera vez el término de Análisis de la varianza (ANOVA) en 1918 (David, 1995). Fisher consideraba el ANOVA como una técnica o procedimiento para analizar las diferencias en el rendimiento de los cultivos en función de las parcelas agrícolas donde se llevaba a cabo el cultivo que diferirían por el tipo de abono o tratamiento que recibían a lo largo de su crecimiento (Gamst, Meyers, y Guarino, 2008).

El ANOVA es una técnica estadística paramétrica que analiza las diferencias entre las puntuaciones medias cuando el diseño utiliza una única variable dependiente que se mide en dos o más grupos o variables independientes o en dos o más mediciones repetidas a los mismos sujetos. Se trata de una técnica de análisis muy popular en las Ciencias Sociales y de la Salud. Es la técnica de análisis de inferencia estadística que se utiliza mayoritariamente en los diseño de investigación univaridos entre-grupos e intra-grupos. Por ejemplo, Keselman y cols. (1998) señalan que en el 93.3% de los artículos que aplican diseños entre-grupos univaridos utilizan la técnica del ANOVA para obtener una solución a sus hipótesis. Como todas las técnicas paramétricas, antes de ejecutar el ANOVA es necesario comprobar los supuestos relacionados con las puntuaciones en la variable de resultados o variable dependiente: independencia de las puntuaciones de los diferentes grupos, normalidad de las puntuaciones y homogeneidad de las varianzas de los grupos junto con la medida realizada al menos en escala de intervalo de la variable dependiente. En muy pocas ocasiones los investigadores señalan en sus artículos el cumplimiento de tales supuestos (Skidmore y Thompson, 2013). Sin embargo, hay que tener en cuenta que los estudios de simulación señalan que la razón F es robusta a leves desviaciones de la normalidad (Harwell, Rubinstein, Hays, y Olds, 1992) y que la razón F es relativamente insensible a la violación de la asunción de normalidad cuando el diseño tiene grupos con el mismo tamaño de observaciones (Glass, Peckham, y Sanders, 1972; Lix, Keselman, y Keselman, 1996).

Valor p (p-value).

En el procedimiento clásico de comprobación de la significación de la hipótesis nula (NHST), el valor p calculado del estadístico computado en el estudio realizado (probabilidad calculada del valor del estadístico, asumiendo que la hipótesis nula es cierta) es: la probabilidad del resultado del hallazgo observado, o un resultado más extremo, cuando la hipótesis nula (H_0) sobre la cuestión que se estudia es verdadera (*The p value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H_0) of a study question is true*). Con otras palabras, el valor p calculado es la probabilidad de los resultados de la muestra (no de la población), o más extremos, asumiendo que la muestra procede de una población donde la hipótesis nula es verdadera. Los

valores de los parámetros poblacionales que se presuponen para calcular el valor de p pueden ser cualquiera y no siempre $H_0 = 0$ (nil-hypothesis). Sin embargo, la mayoría de los paquetes estadísticos (ejemplo, SPSS, JASP, JAMOVI, STATS...) (y consecuentemente, la mayoría de los investigadores) asumen que H_0 de no diferencias entre las puntuaciones (diferencia de cero entre las puntuaciones) es cierto en la población (Cohen, 1994; Thompson, 1996). Generalmente, se considera que la significación estadística o valor p del resultado de la muestra es estadísticamente significativo cuando $p \leq \alpha$, es decir se concluye que hubo un efecto sistemático. Conviene tener muy presente, ya que es un error frecuente, que la significación o valor p no informa de la significación práctica o utilidad del hallazgo (solo el juicio del experto ofrece esa valoración), ni de la significación del tamaño del efecto (esta información requiere computar un estadístico del tamaño del efecto) (Kirk, 1996).

Tamaño del efecto (effect size).

El tamaño del efecto es un estadístico que informa de la magnitud o la fuerza del hallazgo y la dirección de las diferencias entre las puntuaciones de los grupos o de la relación entre las variables. Un tamaño del efecto puede ser una diferencia entre dos medias, un porcentaje o una correlación (Vacha-Hasse y Thompson, 2004). El tamaño del efecto proporciona información sobre cómo de grande es la diferencia entre las puntuaciones de los grupos (ejemplo, d de Cohen) o cómo de fuerte es la relación entre las variables (por ejemplo, r del coeficiente de correlación) o cuanta varianza explica la variable independiente respecto a la varianza total (eta cuadrado, η^2). La información que ofrece el tamaño del efecto no se puede obtener únicamente interpretando el valor p obtenido en el estudio (Volker, 2006) porque el valor p no es un estadístico del tamaño del efecto. Además, no existe siempre una relación directa entre un valor p y la magnitud del efecto, pues un valor p pequeño puede relacionarse con un efecto bajo, medio o alto. Del mismo modo que valor p grande puede relacionarse con un efecto bajo, medio o alto. Tampoco existe siempre una relación directa entre la magnitud del efecto y su valor práctico o clínico. Dependiendo de las circunstancias, un efecto de menor magnitud puede ser más importante que un efecto de mayor magnitud (Durlak, 2009).

Error de Tipo I (Type I error).

El error de Tipo I es la probabilidad de rechazar la hipótesis nula cuando realmente es cierta. Se conoce como criterio de significación o alfa. En un contraste de hipótesis es el rechazo incorrecto de la hipótesis nula. Es decir, la hipótesis nula es falsa y se mantiene tras finalizar el contraste estadístico. Generalmente, el criterio de significación o alfa máximo se sitúa en .05, de tal manera que la hipótesis nula se rechaza cuando $p \leq .05$ y se mantiene cuando $p > .05$.

Nivel de confianza.

Es el valor complementario al valor de alfa (1 - alfa). Se trata de una decisión correcta. El nivel de confianza es la probabilidad de mantener la hipótesis nula siendo realmente cierta. Es decir, la hipótesis nula es cierta y se mantiene correctamente tras finalizar el contraste estadístico.

Error de Tipo II (Type II error).

La probabilidad de mantener la hipótesis nula cuando realmente es falsa. Se conoce como error beta. En un contraste de hipótesis se produce cuando se mantiene de forma incorrecta la hipótesis nula. Es decir, la hipótesis nula es cierta y se rechaza tras finalizar el contraste estadístico.

Potencia estadística (statistical power).

La probabilidad de rechazar correctamente la hipótesis nula (Cohen 1988). Es decir, la hipótesis nula es falsa y se rechaza correctamente tras finalizar el contraste estadístico. Se conoce como potencia estadística del estadístico y representa 1 - error beta.

La potencia estadística depende: 1) del tamaño de la muestra, 2) del tamaño del efecto y 3) del criterio de significación o nivel de confianza, alfa fijado a priori.

En un análisis de potencia a priori (planificando la potencia estadística del diseño del estudio que se va a realizar), el investigador puede calcular el tamaño de muestra necesario (número de observaciones necesarias) para un determinado efecto estimado de un valor concreto y con un nivel de significación o alfa concreto fijado a priori y una potencia estadística deseada (generalmente, al menos de .8) (Cohen, 1988). En general, el valor mínimo recomendado de potencia estadística (1- beta) de .8 se basa en la idea de que con un criterio de significación o alfa de

.05, la ratio del error de Tipo II respecto al error de Tipo I ($.20 / .05 = 4$) es cuatro veces más serio concluir que hay efecto cuando realmente no existe en la población (error de Tipo I) que concluir que no hay efecto cuando realmente sí existe el efecto en la población (error de Tipo II).

Revisión sistemática y meta-análisis.

La revisión sistemática es un método de síntesis de la evidencia o resultados aportados en la literatura sobre una determinada temática común. Requiere un protocolo de revisión con criterios de inclusión y exclusión y una revisión exhaustiva, objetiva y replicable de las publicaciones e informes que existan sobre el constructo objeto de estudio. Generalmente se utilizan bases de datos que deben ser descritas junto con las palabras clave y los criterios de búsqueda en la redacción del informe o artículo. El estudio de meta-análisis se realiza después de la revisión sistemática e implica resumir de forma cuantitativa la evidencia o los resultados con un estadístico del tamaño del efecto medio y su intervalo de confianza. Implica calcular los tamaños del efecto del conjunto de trabajos que forman la revisión sistemática utilizando el mismo estadístico de tamaño del efecto para llevar a cabo finalmente el cómputo del tamaño del efecto medio y su intervalo de confianza. El estudio de meta-análisis incluye analizar la posible heterogeneidad de los estudios y si está presente llevar a cabo un modelo de análisis de efectos aleatorios y un estudio del tamaño del efecto a través de las diferentes variables moderadoras que se propongan como relevantes teóricamente por el investigador.

Referencias

David, H. A. (1995). First (?) occurrence of common terms in mathematical statistics. *American Statistician*, *49*, 121-133. doi: 10.2307/2684625

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*, 917-928.

Gamst, G., Meyers, L. S., & Guarino, A. J. (2008). *Analysis of variance designs: A conceptual and computational approach with SPSS and SAS*. Cambridge: Cambridge University Press.

Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumption underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, *42*, 237-288.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics*, *17*, 315-339. doi:10.2307/1165127

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350-386.

Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746-759.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "F" test. *Review of Educational Research*, *66*, 579-619.

Skidmore, S. T. & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavioral Research*, *45*, 536-546. doi: 10.3758/s13428-012-0257-2

Thompson, B. (1996) AERA Editorial Policies regarding Statistical Significance Testing: Three Suggested Reforms. *Educational Researcher*, *25*, 26-30.

Volker, M. A. (2006). Reporting effect sizes in school psychology research. *Psychology in the Schools, 43*, 653-672.

Anexo 3. Tablas estadísticas

On-line

-Calcular el valor teórico de la Razón F : http://davidmlane.com/hyperstat/F_table.html

1) Distribución F : alfa = .05

Tabla I. Distribución F ($\alpha = .05$, $gl_{entre} = columnas$, $gl_{error} = filas$)

| gl | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 24 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 161.448 | 199.500 | 215.707 | 224.583 | 230.162 | 233.986 | 236.768 | 238.883 | 240.543 | 241.882 | 243.906 | 249.052 |
| 2 | 18.513 | 19.000 | 19.164 | 19.247 | 19.296 | 19.330 | 19.353 | 19.371 | 19.385 | 19.396 | 19.413 | 19.454 |
| 3 | 10.128 | 9.552 | 9.277 | 9.117 | 9.013 | 8.941 | 8.887 | 8.845 | 8.812 | 8.786 | 8.745 | 8.639 |
| 4 | 7.709 | 6.944 | 6.591 | 6.388 | 6.256 | 6.163 | 6.094 | 6.041 | 5.999 | 5.964 | 5.912 | 5.774 |
| 5 | 6.608 | 5.786 | 5.409 | 5.192 | 5.050 | 4.950 | 4.876 | 4.818 | 4.772 | 4.735 | 4.678 | 4.527 |
| 6 | 5.987 | 5.143 | 4.757 | 4.534 | 4.387 | 4.284 | 4.207 | 4.147 | 4.099 | 4.060 | 4.000 | 3.841 |
| 7 | 5.591 | 4.737 | 4.347 | 4.120 | 3.972 | 3.866 | 3.787 | 3.726 | 3.677 | 3.637 | 3.575 | 3.410 |
| 8 | 5.318 | 4.459 | 4.066 | 3.838 | 3.687 | 3.581 | 3.500 | 3.438 | 3.388 | 3.347 | 3.284 | 3.115 |
| 9 | 5.117 | 4.256 | 3.863 | 3.633 | 3.482 | 3.374 | 3.293 | 3.230 | 3.179 | 3.137 | 3.073 | 2.900 |
| 10 | 4.965 | 4.103 | 3.708 | 3.478 | 3.326 | 3.217 | 3.135 | 3.072 | 3.020 | 2.978 | 2.913 | 2.737 |
| 11 | 4.844 | 3.982 | 3.587 | 3.357 | 3.204 | 3.095 | 3.012 | 2.948 | 2.896 | 2.854 | 2.788 | 2.609 |
| 12 | 4.747 | 3.885 | 3.490 | 3.259 | 3.106 | 2.996 | 2.913 | 2.849 | 2.796 | 2.753 | 2.687 | 2.505 |
| 13 | 4.667 | 3.806 | 3.411 | 3.179 | 3.025 | 2.915 | 2.832 | 2.767 | 2.714 | 2.671 | 2.604 | 2.420 |
| 14 | 4.600 | 3.739 | 3.344 | 3.112 | 2.958 | 2.848 | 2.764 | 2.699 | 2.646 | 2.602 | 2.534 | 2.349 |
| 15 | 4.543 | 3.682 | 3.287 | 3.056 | 2.901 | 2.790 | 2.707 | 2.641 | 2.588 | 2.544 | 2.475 | 2.288 |
| 16 | 4.494 | 3.634 | 3.239 | 3.007 | 2.852 | 2.741 | 2.657 | 2.591 | 2.538 | 2.494 | 2.425 | 2.235 |
| 17 | 4.451 | 3.592 | 3.197 | 2.965 | 2.810 | 2.699 | 2.614 | 2.548 | 2.494 | 2.450 | 2.381 | 2.190 |
| 18 | 4.414 | 3.555 | 3.160 | 2.928 | 2.773 | 2.661 | 2.577 | 2.510 | 2.456 | 2.412 | 2.342 | 2.150 |
| 19 | 4.381 | 3.522 | 3.127 | 2.895 | 2.740 | 2.628 | 2.544 | 2.477 | 2.423 | 2.378 | 2.308 | 2.114 |
| 20 | 4.351 | 3.493 | 3.098 | 2.866 | 2.711 | 2.599 | 2.514 | 2.447 | 2.393 | 2.348 | 2.278 | 2.082 |
| 21 | 4.325 | 3.467 | 3.072 | 2.840 | 2.685 | 2.573 | 2.488 | 2.420 | 2.366 | 2.321 | 2.250 | 2.054 |
| 22 | 4.301 | 3.443 | 3.049 | 2.817 | 2.661 | 2.549 | 2.464 | 2.397 | 2.342 | 2.297 | 2.226 | 2.028 |
| 23 | 4.279 | 3.422 | 3.028 | 2.796 | 2.640 | 2.528 | 2.442 | 2.375 | 2.320 | 2.275 | 2.204 | 2.005 |
| 24 | 4.260 | 3.403 | 3.009 | 2.776 | 2.621 | 2.508 | 2.423 | 2.355 | 2.300 | 2.255 | 2.183 | 1.984 |
| 25 | 4.242 | 3.385 | 2.991 | 2.759 | 2.603 | 2.490 | 2.405 | 2.337 | 2.282 | 2.236 | 2.165 | 1.964 |
| 26 | 4.225 | 3.369 | 2.975 | 2.743 | 2.587 | 2.474 | 2.388 | 2.321 | 2.265 | 2.220 | 2.148 | 1.946 |
| 27 | 4.210 | 3.354 | 2.960 | 2.728 | 2.572 | 2.459 | 2.373 | 2.305 | 2.250 | 2.204 | 2.132 | 1.930 |
| 28 | 4.196 | 3.340 | 2.947 | 2.714 | 2.558 | 2.445 | 2.359 | 2.291 | 2.236 | 2.190 | 2.118 | 1.915 |
| 29 | 4.183 | 3.328 | 2.934 | 2.701 | 2.545 | 2.432 | 2.346 | 2.278 | 2.223 | 2.177 | 2.104 | 1.901 |
| 30 | 4.171 | 3.316 | 2.922 | 2.690 | 2.534 | 2.421 | 2.334 | 2.266 | 2.211 | 2.165 | 2.092 | 1.887 |
| 31 | 4.160 | 3.305 | 2.911 | 2.679 | 2.523 | 2.409 | 2.323 | 2.255 | 2.199 | 2.153 | 2.080 | 1.875 |
| 32 | 4.149 | 3.295 | 2.901 | 2.668 | 2.512 | 2.399 | 2.313 | 2.244 | 2.189 | 2.142 | 2.070 | 1.864 |
| 33 | 4.139 | 3.285 | 2.892 | 2.659 | 2.503 | 2.389 | 2.303 | 2.235 | 2.179 | 2.133 | 2.060 | 1.853 |
| 34 | 4.130 | 3.276 | 2.883 | 2.650 | 2.494 | 2.380 | 2.294 | 2.225 | 2.170 | 2.123 | 2.050 | 1.843 |
| 35 | 4.121 | 3.267 | 2.874 | 2.641 | 2.485 | 2.372 | 2.285 | 2.217 | 2.161 | 2.114 | 2.041 | 1.833 |
| 40 | 4.085 | 3.232 | 2.839 | 2.606 | 2.449 | 2.336 | 2.249 | 2.180 | 2.124 | 2.077 | 2.003 | 1.793 |
| 50 | 4.034 | 3.183 | 2.790 | 2.557 | 2.400 | 2.286 | 2.199 | 2.130 | 2.073 | 2.026 | 1.952 | 1.737 |
| 60 | 4.001 | 3.150 | 2.758 | 2.525 | 2.368 | 2.254 | 2.167 | 2.097 | 2.040 | 1.993 | 1.917 | 1.700 |
| 70 | 3.978 | 3.128 | 2.736 | 2.503 | 2.346 | 2.231 | 2.143 | 2.074 | 2.017 | 1.969 | 1.893 | 1.674 |
| 80 | 3.960 | 3.111 | 2.719 | 2.486 | 2.329 | 2.214 | 2.126 | 2.056 | 1.999 | 1.951 | 1.875 | 1.654 |
| 90 | 3.947 | 3.098 | 2.706 | 2.473 | 2.316 | 2.201 | 2.113 | 2.043 | 1.986 | 1.938 | 1.861 | 1.639 |
| 100 | 3.936 | 3.087 | 2.696 | 2.463 | 2.305 | 2.191 | 2.103 | 2.032 | 1.975 | 1.927 | 1.850 | 1.627 |
| 120 | 3.920 | 3.072 | 2.680 | 2.447 | 2.290 | 2.175 | 2.087 | 2.016 | 1.959 | 1.910 | 1.834 | 1.608 |
| 500 | 3.860 | 3.014 | 2.623 | 2.390 | 2.232 | 2.117 | 2.028 | 1.957 | 1.899 | 1.850 | 1.772 | 1.539 |
| 1000 | 3.851 | 3.005 | 2.614 | 2.381 | 2.223 | 2.108 | 2.019 | 1.948 | 1.889 | 1.840 | 1.762 | 1.528 |
| 5000 | 3.843 | 2.998 | 2.607 | 2.374 | 2.216 | 2.100 | 2.011 | 1.940 | 1.882 | 1.833 | 1.754 | 1.519 |

2) Distribución F: alfa = .01

Tabla II. **Distribución F** ($\alpha = .01$, $gl_{entre} = \text{columnas}$, $gl_{error} = \text{filas}$)

| <i>gl</i> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 24 |
|-----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 4052.18 | 4999.50 | 5403.35 | 5624.58 | 5763.65 | 5858.98 | 5928.35 | 5981.08 | 6022.47 | 6055.84 | 6106.32 | 6234.63 |
| 2 | 98.503 | 99.000 | 99.166 | 99.249 | 99.299 | 99.333 | 99.356 | 99.374 | 99.388 | 99.399 | 99.416 | 99.458 |
| 3 | 34.116 | 30.817 | 29.457 | 28.710 | 28.237 | 27.911 | 27.672 | 27.489 | 27.345 | 27.229 | 27.052 | 26.598 |
| 4 | 21.198 | 18.000 | 16.694 | 15.977 | 15.522 | 15.207 | 14.976 | 14.799 | 14.659 | 14.546 | 14.374 | 13.929 |
| 5 | 16.258 | 13.274 | 12.060 | 11.392 | 10.967 | 10.672 | 10.456 | 10.289 | 10.158 | 10.051 | 9.888 | 9.466 |
| 6 | 13.745 | 10.925 | 9.780 | 9.148 | 8.746 | 8.466 | 8.260 | 8.102 | 7.976 | 7.874 | 7.718 | 7.313 |
| 7 | 12.246 | 9.547 | 8.451 | 7.847 | 7.460 | 7.191 | 6.993 | 6.840 | 6.719 | 6.620 | 6.469 | 6.074 |
| 8 | 11.259 | 8.649 | 7.591 | 7.006 | 6.632 | 6.371 | 6.178 | 6.029 | 5.911 | 5.814 | 5.667 | 5.279 |
| 9 | 10.561 | 8.022 | 6.992 | 6.422 | 6.057 | 5.802 | 5.613 | 5.467 | 5.351 | 5.257 | 5.111 | 4.729 |
| 10 | 10.044 | 7.559 | 6.552 | 5.994 | 5.636 | 5.386 | 5.200 | 5.057 | 4.942 | 4.849 | 4.706 | 4.327 |
| 11 | 9.646 | 7.206 | 6.217 | 5.668 | 5.316 | 5.069 | 4.886 | 4.744 | 4.632 | 4.539 | 4.397 | 4.021 |
| 12 | 9.330 | 6.927 | 5.953 | 5.412 | 5.064 | 4.821 | 4.640 | 4.499 | 4.388 | 4.296 | 4.155 | 3.780 |
| 13 | 9.074 | 6.701 | 5.739 | 5.205 | 4.862 | 4.620 | 4.441 | 4.302 | 4.191 | 4.100 | 3.960 | 3.587 |
| 14 | 8.862 | 6.515 | 5.564 | 5.035 | 4.695 | 4.456 | 4.278 | 4.140 | 4.030 | 3.939 | 3.800 | 3.427 |
| 15 | 8.683 | 6.359 | 5.417 | 4.893 | 4.556 | 4.318 | 4.142 | 4.004 | 3.895 | 3.805 | 3.666 | 3.294 |
| 16 | 8.531 | 6.226 | 5.292 | 4.773 | 4.437 | 4.202 | 4.026 | 3.890 | 3.780 | 3.691 | 3.553 | 3.181 |
| 17 | 8.400 | 6.112 | 5.185 | 4.669 | 4.336 | 4.102 | 3.927 | 3.791 | 3.682 | 3.593 | 3.455 | 3.084 |
| 18 | 8.285 | 6.013 | 5.092 | 4.579 | 4.248 | 4.015 | 3.841 | 3.705 | 3.597 | 3.508 | 3.371 | 2.999 |
| 19 | 8.185 | 5.926 | 5.010 | 4.500 | 4.171 | 3.939 | 3.765 | 3.631 | 3.523 | 3.434 | 3.297 | 2.925 |
| 20 | 8.096 | 5.849 | 4.938 | 4.431 | 4.103 | 3.871 | 3.699 | 3.564 | 3.457 | 3.368 | 3.231 | 2.859 |
| 21 | 8.017 | 5.780 | 4.874 | 4.369 | 4.042 | 3.812 | 3.640 | 3.506 | 3.398 | 3.310 | 3.173 | 2.801 |
| 22 | 7.945 | 5.719 | 4.817 | 4.313 | 3.988 | 3.758 | 3.587 | 3.453 | 3.346 | 3.258 | 3.121 | 2.749 |
| 23 | 7.881 | 5.664 | 4.765 | 4.264 | 3.939 | 3.710 | 3.539 | 3.406 | 3.299 | 3.211 | 3.074 | 2.702 |
| 24 | 7.823 | 5.614 | 4.718 | 4.218 | 3.895 | 3.667 | 3.496 | 3.363 | 3.256 | 3.168 | 3.032 | 2.659 |
| 25 | 7.770 | 5.568 | 4.675 | 4.177 | 3.855 | 3.627 | 3.457 | 3.324 | 3.217 | 3.129 | 2.993 | 2.620 |
| 26 | 7.721 | 5.526 | 4.637 | 4.140 | 3.818 | 3.591 | 3.421 | 3.288 | 3.182 | 3.094 | 2.958 | 2.585 |
| 27 | 7.677 | 5.488 | 4.601 | 4.106 | 3.785 | 3.558 | 3.388 | 3.256 | 3.149 | 3.062 | 2.926 | 2.552 |
| 28 | 7.636 | 5.453 | 4.568 | 4.074 | 3.754 | 3.528 | 3.358 | 3.226 | 3.120 | 3.032 | 2.896 | 2.522 |
| 29 | 7.598 | 5.420 | 4.538 | 4.045 | 3.725 | 3.499 | 3.330 | 3.198 | 3.092 | 3.005 | 2.868 | 2.495 |
| 30 | 7.562 | 5.390 | 4.510 | 4.018 | 3.699 | 3.473 | 3.304 | 3.173 | 3.067 | 2.979 | 2.843 | 2.469 |
| 31 | 7.530 | 5.362 | 4.484 | 3.993 | 3.675 | 3.449 | 3.281 | 3.149 | 3.043 | 2.955 | 2.820 | 2.445 |
| 32 | 7.499 | 5.336 | 4.459 | 3.969 | 3.652 | 3.427 | 3.258 | 3.127 | 3.021 | 2.934 | 2.798 | 2.423 |
| 33 | 7.471 | 5.312 | 4.437 | 3.948 | 3.630 | 3.406 | 3.238 | 3.106 | 3.000 | 2.913 | 2.777 | 2.402 |
| 34 | 7.444 | 5.289 | 4.416 | 3.927 | 3.611 | 3.386 | 3.218 | 3.087 | 2.981 | 2.894 | 2.758 | 2.383 |
| 35 | 7.419 | 5.268 | 4.396 | 3.908 | 3.592 | 3.368 | 3.200 | 3.069 | 2.963 | 2.876 | 2.740 | 2.364 |
| 40 | 7.314 | 5.179 | 4.313 | 3.828 | 3.514 | 3.291 | 3.124 | 2.993 | 2.888 | 2.801 | 2.665 | 2.288 |
| 50 | 7.171 | 5.057 | 4.199 | 3.720 | 3.408 | 3.186 | 3.020 | 2.890 | 2.785 | 2.698 | 2.562 | 2.183 |
| 60 | 7.077 | 4.977 | 4.126 | 3.649 | 3.339 | 3.119 | 2.953 | 2.823 | 2.718 | 2.632 | 2.496 | 2.115 |
| 70 | 7.011 | 4.922 | 4.074 | 3.600 | 3.291 | 3.071 | 2.906 | 2.777 | 2.672 | 2.585 | 2.450 | 2.067 |
| 80 | 6.963 | 4.881 | 4.036 | 3.563 | 3.255 | 3.036 | 2.871 | 2.742 | 2.637 | 2.551 | 2.415 | 2.032 |
| 90 | 6.925 | 4.849 | 4.007 | 3.535 | 3.228 | 3.009 | 2.845 | 2.715 | 2.611 | 2.524 | 2.389 | 2.004 |
| 100 | 6.895 | 4.824 | 3.984 | 3.513 | 3.206 | 2.988 | 2.823 | 2.694 | 2.590 | 2.503 | 2.368 | 1.983 |
| 120 | 6.851 | 4.787 | 3.949 | 3.480 | 3.174 | 2.956 | 2.792 | 2.663 | 2.559 | 2.472 | 2.336 | 1.950 |
| 500 | 6.686 | 4.648 | 3.821 | 3.357 | 3.054 | 2.838 | 2.675 | 2.547 | 2.443 | 2.356 | 2.220 | 1.829 |
| 1000 | 6.660 | 4.626 | 3.801 | 3.338 | 3.036 | 2.820 | 2.657 | 2.529 | 2.425 | 2.339 | 2.203 | 1.810 |
| 5000 | 6.640 | 4.609 | 3.786 | 3.323 | 3.021 | 2.806 | 2.643 | 2.515 | 2.411 | 2.324 | 2.188 | 1.795 |

3) Distribución F : alfa = .001

Tabla III. Distribución F ($\alpha = .001$, $gl_{entre} = columnas$, $gl_{error} = filas$)

| gl | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 24 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 405285. | 500001. | 540379. | 562500. | 576408. | 585970. | 592873. | 598153. | 602284. | 605621. | 610669. | 624005. |
| 2 | 998.500 | 999.000 | 999.167 | 999.250 | 999.300 | 999.333 | 999.357 | 999.375 | 999.389 | 999.400 | 999.417 | 999.458 |
| 3 | 167.029 | 148.500 | 141.108 | 137.100 | 134.580 | 132.847 | 131.583 | 130.619 | 129.860 | 129.247 | 128.316 | 125.935 |
| 4 | 74.137 | 61.246 | 56.177 | 53.436 | 51.712 | 50.525 | 49.658 | 48.996 | 48.475 | 48.053 | 47.412 | 45.766 |
| 5 | 47.181 | 37.122 | 33.202 | 31.085 | 29.752 | 28.834 | 28.163 | 27.649 | 27.244 | 26.917 | 26.418 | 25.133 |
| 6 | 35.507 | 27.000 | 23.703 | 21.924 | 20.803 | 20.030 | 19.463 | 19.030 | 18.688 | 18.411 | 17.989 | 16.897 |
| 7 | 29.245 | 21.689 | 18.772 | 17.198 | 16.206 | 15.521 | 15.019 | 14.634 | 14.330 | 14.083 | 13.707 | 12.732 |
| 8 | 25.415 | 18.494 | 15.829 | 14.392 | 13.485 | 12.858 | 12.398 | 12.046 | 11.767 | 11.540 | 11.194 | 10.295 |
| 9 | 22.857 | 16.387 | 13.902 | 12.560 | 11.714 | 11.128 | 10.698 | 10.368 | 10.107 | 9.894 | 9.570 | 8.724 |
| 10 | 21.040 | 14.905 | 12.553 | 11.283 | 10.481 | 9.926 | 9.517 | 9.204 | 8.956 | 8.754 | 8.445 | 7.638 |
| 11 | 19.687 | 13.812 | 11.561 | 10.346 | 9.578 | 9.047 | 8.655 | 8.355 | 8.116 | 7.922 | 7.626 | 6.847 |
| 12 | 18.643 | 12.974 | 10.804 | 9.633 | 8.892 | 8.379 | 8.001 | 7.710 | 7.480 | 7.292 | 7.005 | 6.249 |
| 13 | 17.815 | 12.313 | 10.209 | 9.073 | 8.354 | 7.856 | 7.489 | 7.206 | 6.982 | 6.799 | 6.519 | 5.781 |
| 14 | 17.143 | 11.779 | 9.729 | 8.622 | 7.922 | 7.436 | 7.077 | 6.802 | 6.583 | 6.404 | 6.130 | 5.407 |
| 15 | 16.587 | 11.339 | 9.335 | 8.253 | 7.567 | 7.092 | 6.741 | 6.471 | 6.256 | 6.081 | 5.812 | 5.101 |
| 16 | 16.120 | 10.971 | 9.006 | 7.944 | 7.272 | 6.805 | 6.460 | 6.195 | 5.984 | 5.812 | 5.547 | 4.846 |
| 17 | 15.722 | 10.658 | 8.727 | 7.683 | 7.022 | 6.562 | 6.223 | 5.962 | 5.754 | 5.584 | 5.324 | 4.631 |
| 18 | 15.379 | 10.390 | 8.487 | 7.459 | 6.808 | 6.355 | 6.021 | 5.763 | 5.558 | 5.390 | 5.132 | 4.447 |
| 19 | 15.081 | 10.157 | 8.280 | 7.265 | 6.622 | 6.175 | 5.845 | 5.590 | 5.388 | 5.222 | 4.967 | 4.288 |
| 20 | 14.819 | 9.953 | 8.098 | 7.096 | 6.461 | 6.019 | 5.692 | 5.440 | 5.239 | 5.075 | 4.823 | 4.149 |
| 21 | 14.587 | 9.772 | 7.938 | 6.947 | 6.318 | 5.881 | 5.557 | 5.308 | 5.109 | 4.946 | 4.696 | 4.027 |
| 22 | 14.380 | 9.612 | 7.796 | 6.814 | 6.191 | 5.758 | 5.438 | 5.190 | 4.993 | 4.832 | 4.583 | 3.919 |
| 23 | 14.195 | 9.469 | 7.669 | 6.696 | 6.078 | 5.649 | 5.331 | 5.085 | 4.890 | 4.730 | 4.483 | 3.822 |
| 24 | 14.028 | 9.339 | 7.554 | 6.589 | 5.977 | 5.550 | 5.235 | 4.991 | 4.797 | 4.638 | 4.393 | 3.735 |
| 25 | 13.877 | 9.223 | 7.451 | 6.493 | 5.885 | 5.462 | 5.148 | 4.906 | 4.713 | 4.555 | 4.312 | 3.657 |
| 26 | 13.739 | 9.116 | 7.357 | 6.406 | 5.802 | 5.381 | 5.070 | 4.829 | 4.637 | 4.480 | 4.238 | 3.586 |
| 27 | 13.613 | 9.019 | 7.272 | 6.326 | 5.726 | 5.308 | 4.998 | 4.759 | 4.568 | 4.412 | 4.171 | 3.521 |
| 28 | 13.498 | 8.931 | 7.193 | 6.253 | 5.656 | 5.241 | 4.933 | 4.695 | 4.505 | 4.349 | 4.109 | 3.462 |
| 29 | 13.391 | 8.849 | 7.121 | 6.186 | 5.593 | 5.179 | 4.873 | 4.636 | 4.447 | 4.292 | 4.053 | 3.407 |
| 30 | 13.293 | 8.773 | 7.054 | 6.125 | 5.534 | 5.122 | 4.817 | 4.581 | 4.393 | 4.239 | 4.001 | 3.357 |
| 31 | 13.202 | 8.704 | 6.993 | 6.067 | 5.480 | 5.070 | 4.766 | 4.531 | 4.344 | 4.190 | 3.953 | 3.311 |
| 32 | 13.117 | 8.639 | 6.936 | 6.014 | 5.429 | 5.021 | 4.719 | 4.485 | 4.298 | 4.145 | 3.908 | 3.268 |
| 33 | 13.039 | 8.579 | 6.883 | 5.965 | 5.382 | 4.976 | 4.675 | 4.441 | 4.255 | 4.102 | 3.867 | 3.228 |
| 34 | 12.965 | 8.522 | 6.833 | 5.919 | 5.339 | 4.934 | 4.633 | 4.401 | 4.215 | 4.063 | 3.828 | 3.191 |
| 35 | 12.896 | 8.470 | 6.787 | 5.876 | 5.298 | 4.894 | 4.595 | 4.363 | 4.178 | 4.027 | 3.792 | 3.156 |
| 40 | 12.609 | 8.251 | 6.595 | 5.698 | 5.128 | 4.731 | 4.436 | 4.207 | 4.024 | 3.874 | 3.642 | 3.011 |
| 50 | 12.222 | 7.956 | 6.336 | 5.459 | 4.901 | 4.512 | 4.222 | 3.998 | 3.818 | 3.671 | 3.443 | 2.817 |
| 60 | 11.973 | 7.768 | 6.171 | 5.307 | 4.757 | 4.372 | 4.086 | 3.865 | 3.687 | 3.541 | 3.315 | 2.694 |
| 70 | 11.799 | 7.637 | 6.057 | 5.201 | 4.656 | 4.275 | 3.992 | 3.773 | 3.596 | 3.452 | 3.227 | 2.608 |
| 80 | 11.671 | 7.540 | 5.972 | 5.123 | 4.582 | 4.204 | 3.923 | 3.705 | 3.530 | 3.386 | 3.162 | 2.545 |
| 90 | 11.573 | 7.466 | 5.908 | 5.064 | 4.526 | 4.150 | 3.870 | 3.653 | 3.479 | 3.336 | 3.113 | 2.497 |
| 100 | 11.495 | 7.408 | 5.857 | 5.017 | 4.482 | 4.107 | 3.829 | 3.612 | 3.439 | 3.296 | 3.074 | 2.458 |
| 120 | 11.380 | 7.321 | 5.781 | 4.947 | 4.416 | 4.044 | 3.767 | 3.552 | 3.379 | 3.237 | 3.016 | 2.402 |
| 500 | 10.957 | 7.004 | 5.506 | 4.693 | 4.176 | 3.813 | 3.542 | 3.332 | 3.163 | 3.023 | 2.806 | 2.195 |
| 1000 | 10.892 | 6.956 | 5.464 | 4.655 | 4.139 | 3.778 | 3.508 | 3.299 | 3.130 | 2.991 | 2.774 | 2.164 |
| 5000 | 10.840 | 6.917 | 5.430 | 4.624 | 4.110 | 3.750 | 3.481 | 3.272 | 3.104 | 2.965 | 2.749 | 2.139 |

4) Distribución t de Student

Distribución t de Student

Tabla II. ($T \leq tp, gl$)

| gl | 0.750 | 0.900 | 0.950 | 0.975 | 0.980 | 0.990 | 0.995 | 0.999 |
|------|--------|--------|--------|---------|---------|---------|---------|----------|
| 1 | 1.0000 | 3.0777 | 6.3138 | 12.7062 | 15.8945 | 31.8205 | 63.6567 | 318.3088 |
| 2 | 0.8165 | 1.8856 | 2.9200 | 4.3027 | 4.8487 | 6.9646 | 9.9248 | 22.3271 |
| 3 | 0.7649 | 1.6377 | 2.3534 | 3.1824 | 3.4819 | 4.5407 | 5.8409 | 10.2145 |
| 4 | 0.7407 | 1.5332 | 2.1318 | 2.7764 | 2.9985 | 3.7469 | 4.6041 | 7.1732 |
| 5 | 0.7267 | 1.4759 | 2.0150 | 2.5706 | 2.7565 | 3.3649 | 4.0321 | 5.8934 |
| 6 | 0.7176 | 1.4398 | 1.9432 | 2.4469 | 2.6122 | 3.1427 | 3.7074 | 5.2076 |
| 7 | 0.7111 | 1.4149 | 1.8946 | 2.3646 | 2.5168 | 2.9980 | 3.4995 | 4.7853 |
| 8 | 0.7064 | 1.3968 | 1.8595 | 2.3060 | 2.4490 | 2.8965 | 3.3554 | 4.5008 |
| 9 | 0.7027 | 1.3830 | 1.8331 | 2.2622 | 2.3984 | 2.8214 | 3.2498 | 4.2968 |
| 10 | 0.6998 | 1.3722 | 1.8125 | 2.2281 | 2.3593 | 2.7638 | 3.1693 | 4.1437 |
| 11 | 0.6974 | 1.3634 | 1.7959 | 2.2010 | 2.3281 | 2.7181 | 3.1058 | 4.0247 |
| 12 | 0.6955 | 1.3562 | 1.7823 | 2.1788 | 2.3027 | 2.6810 | 3.0545 | 3.9296 |
| 13 | 0.6938 | 1.3502 | 1.7709 | 2.1604 | 2.2816 | 2.6503 | 3.0123 | 3.8520 |
| 14 | 0.6924 | 1.3450 | 1.7613 | 2.1448 | 2.2638 | 2.6245 | 2.9768 | 3.7874 |
| 15 | 0.6912 | 1.3406 | 1.7531 | 2.1314 | 2.2485 | 2.6025 | 2.9467 | 3.7328 |
| 16 | 0.6901 | 1.3368 | 1.7459 | 2.1199 | 2.2354 | 2.5835 | 2.9208 | 3.6862 |
| 17 | 0.6892 | 1.3334 | 1.7396 | 2.1098 | 2.2238 | 2.5669 | 2.8982 | 3.6458 |
| 18 | 0.6884 | 1.3304 | 1.7341 | 2.1009 | 2.2137 | 2.5524 | 2.8784 | 3.6105 |
| 19 | 0.6876 | 1.3277 | 1.7291 | 2.0930 | 2.2047 | 2.5395 | 2.8609 | 3.5794 |
| 20 | 0.6870 | 1.3253 | 1.7247 | 2.0860 | 2.1967 | 2.5280 | 2.8453 | 3.5518 |
| 21 | 0.6864 | 1.3232 | 1.7207 | 2.0796 | 2.1894 | 2.5176 | 2.8314 | 3.5272 |
| 22 | 0.6858 | 1.3212 | 1.7171 | 2.0739 | 2.1829 | 2.5083 | 2.8188 | 3.5050 |
| 23 | 0.6853 | 1.3195 | 1.7139 | 2.0687 | 2.1770 | 2.4999 | 2.8073 | 3.4850 |
| 24 | 0.6848 | 1.3178 | 1.7109 | 2.0639 | 2.1715 | 2.4922 | 2.7969 | 3.4668 |
| 25 | 0.6844 | 1.3163 | 1.7081 | 2.0595 | 2.1666 | 2.4851 | 2.7874 | 3.4502 |
| 26 | 0.6840 | 1.3150 | 1.7056 | 2.0555 | 2.1620 | 2.4786 | 2.7787 | 3.4350 |
| 27 | 0.6837 | 1.3137 | 1.7033 | 2.0518 | 2.1578 | 2.4727 | 2.7707 | 3.4210 |
| 28 | 0.6834 | 1.3125 | 1.7011 | 2.0484 | 2.1539 | 2.4671 | 2.7633 | 3.4082 |
| 29 | 0.6830 | 1.3114 | 1.6991 | 2.0452 | 2.1503 | 2.4620 | 2.7564 | 3.3962 |
| 30 | 0.6828 | 1.3104 | 1.6973 | 2.0423 | 2.1470 | 2.4573 | 2.7500 | 3.3852 |
| 31 | 0.6825 | 1.3095 | 1.6955 | 2.0395 | 2.1438 | 2.4528 | 2.7440 | 3.3749 |
| 32 | 0.6822 | 1.3086 | 1.6939 | 2.0369 | 2.1409 | 2.4487 | 2.7385 | 3.3653 |
| 33 | 0.6820 | 1.3077 | 1.6924 | 2.0345 | 2.1382 | 2.4448 | 2.7333 | 3.3563 |
| 34 | 0.6818 | 1.3070 | 1.6909 | 2.0322 | 2.1356 | 2.4411 | 2.7284 | 3.3479 |
| 35 | 0.6816 | 1.3062 | 1.6896 | 2.0301 | 2.1332 | 2.4377 | 2.7238 | 3.3400 |
| 36 | 0.6814 | 1.3055 | 1.6883 | 2.0281 | 2.1309 | 2.4345 | 2.7195 | 3.3326 |
| 37 | 0.6812 | 1.3049 | 1.6871 | 2.0262 | 2.1287 | 2.4314 | 2.7154 | 3.3256 |
| 38 | 0.6810 | 1.3042 | 1.6860 | 2.0244 | 2.1267 | 2.4286 | 2.7116 | 3.3190 |
| 39 | 0.6808 | 1.3036 | 1.6849 | 2.0227 | 2.1247 | 2.4258 | 2.7079 | 3.3128 |
| 40 | 0.6807 | 1.3031 | 1.6839 | 2.0211 | 2.1229 | 2.4233 | 2.7045 | 3.3069 |
| 45 | 0.6800 | 1.3006 | 1.6794 | 2.0141 | 2.1150 | 2.4121 | 2.6896 | 3.2815 |
| 50 | 0.6794 | 1.2987 | 1.6759 | 2.0086 | 2.1087 | 2.4033 | 2.6778 | 3.2614 |
| 60 | 0.6786 | 1.2958 | 1.6706 | 2.0003 | 2.0994 | 2.3901 | 2.6603 | 3.2317 |
| 100 | 0.6770 | 1.2901 | 1.6602 | 1.9840 | 2.0809 | 2.3642 | 2.6259 | 3.1737 |
| 120 | 0.6765 | 1.2886 | 1.6577 | 1.9799 | 2.0763 | 2.3578 | 2.6174 | 3.1595 |
| 500 | 0.6750 | 1.2832 | 1.6479 | 1.9647 | 2.0591 | 2.3338 | 2.5857 | 3.1066 |
| 1000 | 0.6747 | 1.2824 | 1.6464 | 1.9623 | 2.0564 | 2.3301 | 2.5808 | 3.0984 |
| 5000 | 0.6745 | 1.2817 | 1.6452 | 1.9604 | 2.0543 | 2.3271 | 2.5768 | 3.0919 |

Anexo 4. Comparación entre diferentes pruebas: tamaño del efecto pequeño, mediano y grande

On-line

-Calcular el tamaño del efecto y su intervalo de confianza:

<https://campbellcollaboration.org/research-resources/effect-size-calculator.html>

Tabla 1. Pruebas estadísticas y valores de los tamaños del efecto

| Prueba | Tamaño del efecto | Pequeño | Mediano | Grande |
|---|-------------------|---------|---------|--------|
| Diferencia de medias estandarizada, d de Cohen | d | 0.20 | 0.50 | 0.80 |
| Correlación de Pearson | r | 0.10 | 0.30 | 0.50 |
| Correlación de Pearson al cuadrado | r^2 | 0.01 | 0.09 | 0.25 |
| Correlación biserial-puntual | r_{bp} | 0.10 | 0.24 | 0.37 |
| Correlación biserial-puntual al cuadrado | r_{bp}^2 | 0.01 | 0.06 | 0.14 |
| Eta Cuadrado | η^2 | 0.01 | 0.06 | 0.14 |
| Coefficiente de determinación | R^2 | 0.01 | 0.06 | 0.14 |
| Omega Cuadrado | ω^2 | 0.01 | 0.06 | 0.14 |
| f de Cohen del ANOVA unifactorial | f | 0.10 | 0.25 | 0.40 |
| f^2 del análisis de regresión (más de dos grupos) | f^2 | 0.02 | 0.15 | 0.35 |
| Ji Cuadrado | w | 0.10 | 0.30 | 0.50 |
| Odds Ratio | OR | 1.29 | 1.88 | 2.69 |
| Porcentaje de solapamiento | OL % | 85.3 | 66.6 | 52.6 |
| Porcentaje de no solapamiento (1-% de solapamiento) (U1 de Cohen) | U1 | 14.7% | 33.4% | 47.4% |
| Common Language Effect size | CL (AUC) | 0.56 | 0.64 | 0.71 |

Anexo 5. Autoevaluaciones

A continuación se ha elaborado un cuestionario de autoevaluación cuyas respuestas se encuentran en la web: <https://www.uv.es/friasnav> en el apartado que se ha desarrollado para el libro:

Frías-Navarro, D. y Pascual-Soler, M. (Eds.) (2020). *Diseño de la investigación, análisis y presentación de resultados*. Universidad de Valencia. España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> Disponible en: osf.io/kngtp

1) Autoevaluación 1

Instrucciones

Para realizar la autoevaluación hay que tener en cuenta que solamente se puede dar una respuesta a cada pregunta. Es decir, no se pueden señalar dos o más opciones como respuestas a la pregunta. Si se considera que hay dos respuestas verdaderas entonces hay que seleccionar la opción de “No se puede responder a la pregunta”, ya que solamente se puede seleccionar una opción de respuesta. Si todas las respuestas son falsas y existe la opción de “No se puede responder a la pregunta” entonces dicha opción será la correcta. Del mismo modo, si todas las respuestas son verdaderas y existe la opción de “No se puede responder a la pregunta” entonces dicha opción será la correcta.

Las respuestas correctas a las autoevaluaciones se encuentran en un documento que se depositará en la página web docente una vez que el alumando haya realizado las autoevaluaciones. Por lo tanto, se irán corrigiendo parte de las preguntas en clase y la otra parte se dará la opción correcta en la página web docente.

Cuestionario

1. Qué quiere decir maximizar la varianza sistemática primaria:

- A. Aumentar el valor de alfa que se fija a priori ($> .05$) para maximizar la probabilidad de detectar el efecto que se desea estudiar.
- B. Diseñar un estudio con un número amplio de observaciones (> 100) para maximizar el efecto que se desea estudiar.

- C. Planificar en el diseño las condiciones necesarias que maximicen el efecto que se desea estudiar.
 - D. Disminuir el valor de alfa que se fija a priori ($> .05$) para maximizar la probabilidad de detectar el efecto que se desea estudiar.
- 2. En un diseño de investigación, el problema de la tercera variable se refiere a:**
- A. La validez de conclusión estadística.
 - B. La validez interna.
 - C. La validez externa.
 - D. La validez de constructo.
- 3. Los resultados de un diseño tienen validez interna cuando la variabilidad de las puntuaciones en la variable dependiente es:**
- A. Sistemática y provocada por el efecto de la variable independiente.
 - B. Sistemática, aleatoria y provocada por el efecto de la variable medida.
 - C. No sistemática y constante y producida por el experimentador.
 - D. No se puede responder a la pregunta.
- 4. Un investigador desea estudiar la cantidad de sudor que segregan los pacientes fóbicos según el grado de cercanía del estímulo que desencadena el miedo. Por lo tanto, en el diseño:**
- A. La variable dependiente será el tipo de estímulo que desencadena la fobia y la independiente la intensidad de la secreción.
 - B. La variable independiente será la intensidad de la secreción y la dependiente la cercanía del estímulo que desencadene la fobia.
 - C. La variable independiente será la cercanía del estímulo que desencadene la fobia y la variable dependiente la intensidad de la secreción.
 - D. La variable dependiente será el grado de cercanía del estímulo y la independiente la intensidad de la secreción.
- 5. El proceso de búsqueda y localización de la información bibliográfica será más válido cuando:**
- A. El problema de investigación (necesidad de conocimiento) se plantea una vez examinados los resultados del estudio.

- B. El problema de investigación (necesidad de conocimiento) no se define previamente a dicho proceso de búsqueda.
- C. El problema de investigación (necesidad de conocimiento) se define de forma concreta previamente a dicho proceso de búsqueda.
- D. El problema de investigación (necesidad de conocimiento) se define de forma amplia para así poder captar más referencias bibliográficas.

6. En el denominado principio MAXMINCON se señala que hay que controlar:

- A. La asignación aleatoria de las unidades experimentales a los grupos o condiciones de medida.
- B. La varianza no sistemática (aleatoria).
- C. La varianza sistemática secundaria.
- D. La varianza sistemática primaria.

7. La validez interna de un diseño está relacionada con:

- A. La selección aleatoria de la muestra.
- B. La generalización de los resultados.
- C. La aplicación correcta de las pruebas de inferencia estadística.
- D. La minimización del sesgo y el incremento en la precisión de la estimación del efecto.

8. El efecto del tratamiento está vinculado con:

- A. La varianza sistemática primaria.
- B. La varianza sistemática secundaria.
- C. La varianza sistemática terciaria.
- D. La varianza sistemática aleatoria.

9. En un diseño de investigación la manipulación de la variable significa que:

- A. El investigador controla las condiciones de la variable independiente.
- B. El investigador manipula las puntuaciones en la variable dependiente.
- C. El investigador actúa sin ética y/o integridad.
- D. El investigador manipula las opiniones de los participantes para facilitar que cooperen en el desarrollo del estudio.

10. Las variables extrañas en un modelo de diseño de investigación son:

- A. Son una fuente de varianza sistemática primaria.
- B. Variables dependientes que no forman parte de la hipótesis.
- C. Variables independientes que no forman parte de la hipótesis.
- D. Son una fuente de varianza sistemática aleatoria.

2) Autoevaluación 2

SUPÒSIT D'INVESTIGACIÓ 1. En els últims anys s'ha comprovat que les persones majors i fins i tot aquelles diagnosticades de malaltia d'Alzheimer en fase lleu o moderada, encara que de forma limitada, són capaços d'aprendre (Calero, 2000).

Les bases biològiques d'aquesta capacitat d'aprenentatge procedeix de l'àmplia evidència empírica sobre la capacitat de les neurones lesionades per regenerar-se i establir noves connexions (Goldman, 1995).

S'ha demostrat la plasticitat del sistema nerviós o neuroplasticitat en el cervell de l'ancià, fins i tot quan la demència és lleu, però no passa el mateix en les fases greus de la malaltia ja que hi ha una gran pèrdua de neurones i falta de connexions sinàptiques (Kass, 1995).

En el present estudi s'ha dut a terme un disseny on es mesurava a l'inici la línia base del deteriorament cognitiu (es diagnosticava el deteriorament cognitiu), després es realitzava un entrenament en tasques cognitives (psicoestimulació cognitiva) i posteriorment (al cap de sis mesos) es tornava a mesurar els subjectes en aquesta variable per observar si s'havia produït millora en el potencial d'aprenentatge. El deteriorament cognitiu es mesurava amb el Mini-Mental State Examination (MMSE) on a major puntuació major deteriorament.

En l'estudi van participar 8 pacients d'Alzheimer en fase lleu de deteriorament cognitiu (edat mitjana: 75.3 anys, DT: 6.4; 6 homes i 2 dones) diagnosticats per la Unitat de Diagnòstic de l'hospital. Quatre d'ells van participar en un programa de

psicoestimulació cognitiva (PPC) i la resta no participaven en cap programa (grup control).

L'assignació dels subjectes als grups experimental i control es va deure a raons clíniques ja que aquells subjectes que tenien un pitjor estat mental a la línia base i un major deteriorament en memòria aspecte-espacial, àudio verbal i fluïdesa verbal pertanyien al grup experimental.

Tots els pacients (grup experimental i control) portaven un tractament estable més de quatre mesos amb inhibidors d'acetilcolinesterasa abans de començar l'estudi i van continuar amb el mateix tractament durant tot el procés experimental.

Tots els contrastos estadístics són bilaterals (alfa = .05).

- A. El investigador controla las condiciones de la variable independiente.
- B. El investigador manipula las puntuaciones en la variable dependiente.
- C. El investigador actúa sin ética y/o integridad.
- D. El investigador manipula las opiniones de los participantes para facilitar que cooperen en el desarrollo del estudio.

11. La metodologia del disseny de la investigació és:

- A. Experimental.
- B. Quasi-experimental.
- C. No Experimental.
- D. Qualitativa.

12. La o les variables independents en el 'Disseny Tipus A' són:

- A. El grau ansietat del pacient en la fase de línia base (lleu / no lleu).
- B. La variable Grup de subjectes (experimental/control).
- C. El sexe i el grup de subjectes.
- D. Les fases de línia base i la de post-intervenció.

13. Els efectes del entrenament cognitiu:

- A. Es mesuren en la línia base.
- B. Representen a la variable dependent del disseny.
- C. És una variable estranya que es necessari controlar.
- D. Totes las opcions anteriors són verdaderes

14. Una variable estranya que s' ha controlat per constància:

- A. La fase de la malatia de Alzheimer.
- B. El sexe dels pacients.
- C. Tractament farmacològic prescrit.
- D. Les opcions A i C són verdaderes.

15. Teniendo en cuenta los siguientes resultados, se puede concluir que:

Contraste de Levene sobre la igualdad de las varianzas error^a

Variable dependiente: NOPpost

| F | gl1 | gl2 | Sig. |
|-------|-----|-----|------|
| 1,495 | 5 | 302 | ,191 |

Contrasta la hipótesis nula de que la varianza error de la variable dependiente es igual a lo largo de todos los grupos.

a. Diseño: Intersección + NOPpre + Parenting + Clima + Parenting * Clima

- A. Las diferencias entre las puntuaciones medias de un diseño de cinco grupos no son estadísticamente significativas.
- B. Se cumple el supuesto de homocedasticidad o igualdad de las varianzas de los seis grupos del diseño.
- C. Existe heterocedasticidad o falta de homogeneidad entre las varianzas de los cinco grupos del diseño.
- D. Las tres opciones anteriores son falsas.

16. Cuando el investigador no conoce qué variable extraña podría contaminar los resultados la mejor opción es:

- A. La técnica de la constancia de la variable extraña.
- B. La técnica de la eliminación de la variable extraña.
- C. La técnica de la asignación aleatoria de la variable extraña.
- D. La técnica del bloqueo de la variable extraña.

17. Teniendo en cuenta los resultados siguientes, se puede concluir que:

| Comparaciones múltiples | | | | | | |
|-------------------------|-------------|----------------------------|------------|------|----------------------------|-----------------|
| NOPpost DHS de Tukey | | | | | | |
| (I)Clima | (J)Clima | Diferencia de medias (I-J) | Error típ. | Sig. | Intervalo de confianza 95% | |
| | | | | | Límite inferior | Límite superior |
| Luminosidad | Oscuridad | -1,86* | ,741 | ,034 | -3,60 | -,11 |
| | Tiniebla | -,84 | ,698 | ,455 | -2,48 | ,81 |
| Oscuridad | Luminosidad | 1,86* | ,741 | ,034 | ,11 | 3,60 |
| | Tiniebla | 1,02 | ,676 | ,288 | -,57 | 2,61 |
| Tiniebla | Luminosidad | ,84 | ,698 | ,455 | -,81 | 2,48 |
| | Oscuridad | -1,02 | ,676 | ,288 | -2,61 | ,57 |

Subconjuntos homogéneos

NOPpost

DHS de Tukey^{a,b,c}

| Clima | N | Subconjunto | |
|-------------|-----|-------------|-------|
| | | 1 | 2 |
| Luminosidad | 86 | 11,47 | |
| Tiniebla | 126 | 12,30 | 12,30 |
| Oscuridad | 96 | | 13,32 |
| Sig. | | ,463 | ,318 |

- A. No se logra rechazar la hipótesis nula en ninguna diferencia de medias.
- B. Las diferencias estadísticamente significativas solamente se encuentran entre luminosidad y tiniebla, siendo mayor la puntuación en la condición de tiniebla.
- C. Las diferencias estadísticamente significativas solamente se encuentran entre luminosidad y oscuridad, siendo menor la puntuación en la condición de oscuridad.
- D. Las diferencias estadísticamente significativas solamente se encuentran entre luminosidad y oscuridad, siendo mayor la puntuación en la condición de oscuridad.

18. El error de tipo I se puede cometer cuando:

- A. Los resultados obtenidos no son estadísticamente significativos.
- B. El valor p de probabilidad es menor a 0.05.
- C. Se mantiene la hipótesis nula.
- D. El valor de alfa es menor a 0.05.

19. Cometer Error de Tipo II es imposible cuando:

- A. Los resultados obtenidos no son estadísticamente significativos.
- B. El valor p de probabilidad es menor a 0.05.
- C. Se mantiene la hipótesis nula.
- D. El valor de alfa es menor a 0.05.

20. Supongamos que un investigador lleva a cabo un estudio y los resultados del efecto del tratamiento A del análisis de varianza que ofrece el SPSS señalan que $F(5, 123) = 43$, $p = .000$. Qué conclusión habrá obtenido el investigador o investigadora con su análisis:

- A. No se puede rechazar la hipótesis nula porque $p = 0$.
- B. Hay diferencias estadísticamente significativas entre los grupos, pero no se sabe entre qué pares de medias.
- C. Se mantiene la hipótesis nula.
- D. Las opciones A y C son correctas.

21. $F(5, 123) = 43$: ¿cuántos grupos tiene el diseño entre-grupos unifactorial?

- A. 3.
- B. 4.
- C. 5.
- D. 6.

22. Si el diseño entre-grupos unifactorial: $F(5, 123) = 43$, ¿cuántas comparaciones de medias dos a dos simples no redundantes se podrían realizar?

- A. 3.
- B. 15.
- C. 16.
- D. 18.

23. El diseño entre-grupos unifactorial: $F(5, 123) = 43$. Si el investigador decide comparar el último grupo con el resto y realiza solamente cinco comparaciones simples, qué procedimiento de hipótesis específicas sería el más adecuado desde el punto de la validez de conclusión estadística:

- A. Bonferroni.
- B. Tukey.
- C. Dunnett.
- D. Las opciones B y C serían correctas.

24. La técnica de control de la asignación aleatoria implica:

- A. Aumentar la variabilidad de los grupos antes de administrar el tratamiento.
- B. Aumentar el efecto del tratamiento aplicado.
- C. Disminuye el valor del tamaño del efecto y el valor p de probabilidad.
- D. Disminuir la probabilidad de diferencias previas al tratamiento entre los grupos del diseño de investigación.

25. ¿Qué definición de valor p de probabilidad es la adecuada?

- A. El valor de alfa o más extremo asumiendo que la hipótesis nula es verdadera.
- B. La probabilidad del resultado obtenido o más extremo si la hipótesis nula es verdadera.

- C. La probabilidad del resultado obtenido o más extremo si la hipótesis nula es falsa.
- D. La probabilidad de que la hipótesis nula sea falsa dados unos datos o datos más extremos.

26. La potencia estadística se refiere a la:

- A. Probabilidad de mantener la hipótesis nula siendo realmente falsa.
- B. Probabilidad de detectar una diferencia cuando realmente existe.
- C. Probabilidad de que la hipótesis nula sea verdadera.
- D. Las opciones B y C son correctas.

SUPÒSIT D'INVESTIGACIÓ 2. En el treball de Beneyto i García (2012) es pretén contrastar si un tret positiu de personalitat com és l'optimisme podria minimitzar el biaix en el record de paraules després d'induir una emoció determinada. En un nou estudi elaborat per la Universitat de València, es va induir un determinat estat emocional a 12 adolescents homes que es van seleccionar aleatòriament de un Centre Juvenil d' Oci (distribuïts aleatòriament i de forma equilibrada en tres grups en funció de si l'estat emocional induït era positiu (grup 1), negatiu (grup 2) o neutre (grup 3).

En l'experiment, en primer lloc, tots els subjectes visionaven una mateixa llista de 60 paraules amb valència neutra, instant-los a recordar aquestes paraules. Tots els participants van rebre les mateixes instruccions: "aneu a participar en un experiment de memòria. Quan siguis preparat se't presentarà una sèrie de paraules sobre les que versarà la prova". Quan el subjecte ho indicava es donava pas a la presentació amb la llista de paraules.

Un cop finalitzat el visionat de les paraules, se li indicava que anava a veure una pel·lícula curta i que intentés ficar-se al màxim en l'escena. Depenent de la condició experimental a la qual estigués adscrit es va visionar el curt corresponent. Finalitzat el visionat de la pel·lícula, se'ls va lliurar un full en blanc indicant-los que disposaven de 3 minuts per anotar totes les paraules que recordessin de la presentació que prèviament havien vist.

Per a la inducció emocional es van emprar tres curts de pel·lícula extretes de l'estudi de Rottenberg, Ray i Gross (2007). Per induir l'emoció positiva es va emprar el curt "Quan Harry va trobar la Sally", que presenta una situació còmica en una cafeteria amb una durada de 2,357 minuts. Per a la emoció negativa es va emprar

el curt "Crida Llibertat", el qual presenta una matança de tall racista i té una durada de 2,361 minuts. Finalment, per l'emoció neutra es va emprar el curt "Sticks", que és una presentació semblant a un estalvi de pantalla d'ordinador on van apareixent uns bastons a la pantalla; aquesta pel·lícula no provoca cap emoció i va tenir una durada de 2,355 minuts. Per mesurar el nivell d'optimisme es va utilitzar l'instrument LOT-R (Scheier, Carver i Bridges, 1994) en la versió adaptada i validada al castellà de Fernández i Bermudes (1999). Per obtenir els grups experimentals es va dividir als participants en dos grups en virtut de la puntuació obtinguda en el qüestionari LOT-R: alt optimisme amb una mitjana igual o superior a 22 (grup b1) i baix optimisme amb una mitjana inferior a 22 (grup b2). La hipòtesi d'investigació planteja que els subjectes que tenen un estat emocionat positiu recordaran més paraules però, aquest efecte estarà moderat per un grau d'optimisme positiu. Els resultats trobats indiquen una tendència dels més optimistes a recordar major nombre de paraules però si l'estat emocional en què es troben també és positiu ja que si és negatiu ja no es diferencien dels subjectes menys optimistes.

27. Hi ha manipulació de la o les variables independents?

- A. Hi ha dos variables independents manipulades.
- B. Les pel·lícules s'han utilitzat per manipular la variable independent.
- C. La divisió del subjectes en nivell d'optimisme és una manipulació de la variable independent.
- D. Les tres opcions anteriors són correctes.

28. Hi ha assignació aleatòria?

- A. Sí, perquè els participants es van seleccionar aleatòriament
- B. Sí, perquè els participants s'han distribuït aleatòriament.
- C. Sí, perquè els participants s'han distribuït de forma equilibrada.
- D. Les opcions B i C són verdaderes.

29. Quin tipus de metodologia s'ha utilitzat?

- A. Experimental.
- B. Observacional.
- C. Quasi-experimental.
- D. No experimental.

30. La variable dependent és:

- A. El nombre de paraules recordades.
- B. El grau d'optimisme.

- C. La duració de la pel·lícula.
- D. La inducció emocional.

31. Les variables independents són:

- A. El nombre de paraules recordades i la duració de la pel·lícula.
- B. El grau d'optimisme i el nombre de paraules recordades.
- C. La duració de la pel·lícula i el sexe.
- D. La inducció emocional i el grau d'optimisme.

32. Respecte a la metodologia, es tracta d'un disseny:

- A. Totalment aleatori.
- B. No aleatori.
- C. Parcialment aleatori.
- D. No es pot saber amb les dades del supòsit.

33. Què variable o variables s'han controlat per constància

- A. La valència de les paraules de la llista.
- B. La durada de la pel·lícula.
- C. El sexe.
- D. Totes les opcions anteriors son verdaderes.

34. Segons l'hipòtesi, quin grup recordarà més paraules:

- A. Grup a1b1.
- B. Grup a1b2.
- C. Grup a2b1.
- D. Grup a2b2.

SUPÒSIT D'INVESTIGACIÓ 3. El trastorn obsessiu-compulsiu (TOC) és un trastorn heterogeni pel que fa al contingut de les obsessions i compulsions. El rentat, una de les compulsions més freqüents, s'ha associat a diferents variables com: la por al contagi físic o mental, la por a l'afecte negatiu i a la pèrdua de control, un elevat perfeccionisme o l'evitació de la sensació de brutícia o de inacabat. Un grup d'investigadors tenen el propòsit d'estudiar la simptomatologia psicopatològica i la seua resposta al llarg de tres fases d'intervenció cognitiva per tractar el TOC: fase de pretest, fase primera de avaluació i fase de post-test. L'estudi compta amb 20 subjectes, homes, i el tractament té una durada de cinc mesos. També es mesura l'edat. Quant més alta és la puntuació, més psicopatologia tenen els subjectes. El valor de alfa en aquesta investigació es de $\alpha = .01$.

35. Assenyala les variables dependents i independent, i possibles variables estranyes

36. Assenyala la metodologia, *N*, *n*, *S*, tipus de disseny

SUPÒSIT D'INVESTIGACIÓ 4. En el treball de Molero, Navas i Morales (2001) se assenyalen els següents comentaris: "els conceptes de "prejudici manifest" i "prejudici subtil" de Pettigrew i Meertens (1995) impliquen "l'exclusió social del grup objecte de prejudici encara que a través de diferents vies. El prejudici manifest ho fa a través del rebuig directe i sense pal·liatius dels membres de l'exogrup per considerar-los "biològicament inferiors". El prejudici subtil condueix a un rebuig indirecte que es justifica per la defensa dels valors tradicionals que els immigrants qüestionen o no comparteixen, i l'exageració de les diferències culturals entre la societat d'acollida i la d'arribada, entre "nosaltres" i "ells ". Tot això porta a la negació d'emocions positives cap als membres de l'exogrup. Per això la persona amb prejudici subtil no té, o almenys no expressa, emocions negatives cap als immigrants, però és incapaç també de manifestar emocions positives cap a ells. Cal assenyalar que la persona que té prejudici subtil cap a un determinat grup no és conscient del seu prejudici i de les conductes discriminatòries que aquest prejudici pot arribar a produir" (p. 20).

Un grup d'investigadors esta interessat pel tema del prejudici subtil i desitgen analitzar la seua relació amb les emociones positives. Els investigadors construeixen la tipologia de prejudici (igualitari, subtil i fanàtic) i plantegen analitzar la seua relació amb les puntuacions d'una escala de emocions positives.

La mostra esta formada per 12 estudiants de 4º de l' ESO, seleccionats aleatòriament d'un institut, 4 homes i 4 dones. La seua hipòtesi d'investigació planteja que hi haurà un efecte d'interacció les variables.

37. Si un del factors és el sexe (més prejudici en els homes), qui serà el segon factor?.

- A. El tipus de emocions: positives o negatives.
- B. La tipologia de prejudici.
- C. El curs acadèmic.
- D. La intensitat de l'emoció positiva manifestada.

38. La variable dependent és:

- A. El tipus de emocions: positives o negatives.
- B. La tipologia de prejudici.
- C. El curs acadèmic.
- D. La intensitat de l'emoció positiva manifestada.

39. La metodologia de la investigació és:

- A. Experimental amb restriccions en la aleatorització.
- B. Experimental.
- C. Quasi-experimental.
- D. No experimental.

SUPÒSIT D'INVESTIGACIÓ 4. Un equip d'investigació desitja estudiar l'efecte de la intervenció amb pallassos en la avantsala d'operacions sobre l'ansietat dels xiquets respecte a un grup de control que no rep pre-medicació sedant ni cap altre tipus de teràpia i un grup que rep un ansiolític ('midazolam'). Els xiquets s'assignen a la condició d'investigació en funció del criteri del metge i la severitat de la seua malaltia.

Seleccionen a 12 pacients entre 16 anys i s'utilitza un disseny ortogonal o equilibrat.

La simptomatologia de ansietat es mesura amb un instrument tipus termòmetre on el xiquet ha de valorar de 0 a 22 el seu nivell de ansietat. Tots els participants són homes.

La hipòtesi de treball planteja que el nivell d'ansietat dels xiquets del grup experimental que rep la intervenció dels pallassos d'hospital serà diferent al dels xiquets de la resta dels grups.

40. La variable dependent és:

- A. Grups d'intervenció.
- B. Simptomatologia depressiva.
- C. Simptomatologia ansiosa.

41. Hi ha alguna variable estranya pertorbadora?

- A. El sexe.
- B. La gravetat de la malaltia.
- C. La edat.

42. Hi ha alguna variable estranya controlada?

- A. El sexe.
- B. La gravetat de la malaltia.
- C. El metge.

43. La metodologia de la investigació és:

- A. Experimental.
- B. Quasi-experimental.
- C. No experimental.

44. Quines tècniques de control s'han aplicat en el disseny:

- A. Aleatorització.
- B. Constància.
- C. Les opcions de resposta A i B són correctes.

SUPUESTO DE INVESTIGACIÓN 5. Un investigador pretende comprobar la eficacia de un fármaco Z para incrementar la extroversión de los individuos. 12 voluntarios son asignados aleatoriamente a una de las tres condiciones experimentales siguientes. Se supone que el fármaco Z producirá mejores resultados que cualquiera de las otras dos sustancias que actualmente están comercializadas. El objetivo del estudio es comparar el efecto del fármaco Z respecto al fármaco A y el fármaco B. Los resultados del experimento son los siguientes.

| Condición | Puntuaciones |
|-----------|----------------|
| Fármaco A | 26, 22, 25, 23 |
| Fármaco B | 26, 25, 19, 22 |
| Fármaco Z | 35, 30, 32, 35 |

1. Qué procedimiento de análisis procede dada la hipótesis de investigación.
2. Atendiendo únicamente a la tendencia de las medias, ¿se observa el planteamiento de la hipótesis sustantiva?
3. Ejecutar el análisis de la varianza (ANOVA) y aplicar la prueba de contraste de hipótesis específica que sea más adecuada.
4. Realizar un informe de los hallazgos.

SUPUESTO DE INVESTIGACIÓN 6. Supongamos ahora que el investigador o investigadora hubiese planteado analizar el efecto del fármaco Z respecto a la media de los fármacos A y B. ¿Qué procedimiento de contraste de hipótesis específicas hubiese sido más adecuado?

SUPUESTO DE INVESTIGACIÓN 7. El sueño es muy importante en el desarrollo infantil y favorece la maduración. Los patrones electroencefalográficos (EEG) durante el sueño presentan cambios que, si no se tienen en cuenta, pueden confundirse con actividad clínica paroxística o alteraciones bruscas del trazado, aunque son fisiológicos, especialmente en los niños. En general, el EEG es uno de los estudios que suelen realizarse en los niños con problemas de aprendizaje ya que algunas investigaciones señalan que existe actividad paroxística (*hipersincronía hipnagógica*) en los niños con este problema.

Un investigador está interesado por el estudio de la relación entre la hipersincronía hipnagógica y las dificultades del aprendizaje. En un primer momento de su investigación planteó el siguiente estudio. Considerando que existe una relación lineal directa entre las variables objeto de estudio, de tal manera que mayor hipersincronía mayores son los problemas de aprendizaje, selecciona nueve niños de cinco años. Los nueve sujetos habían sido diagnosticados en sus Centros Escolares con un problema de trastornos del aprendizaje antes de comenzar el estudio. Aleatoriamente tres de ellos no recibieron ningún fármaco actuando como grupo de control (a_1), otros tres recibieron asistencia psicológica para su trastorno escolar (a_2) y el resto de niños recibió un fármaco que provocaba la relajación neuronal (a_3).

Una vez finalizado el tratamiento el investigador midió el nivel electroencefalográfico computando el número de alteraciones bruscas del trazado o cambios paroxísticos producidos.

Su hipótesis señala que los niños que no son sometidos a ningún tipo de intervención tendrán un mayor número de alteraciones en el EEG, siendo el número menor cuando reciben tratamiento psicológico.

Los resultados de parte del análisis de la varianza y de la estimación de los efectos se presentan a continuación.

Ejercicios: completa los resultados, ejecuta el ANOVA y redacta los resultados del informe de investigación.

| Fuente | SC | gl | MC | Razón F | p |
|--------------|----|----|----|---------|------|
| <i>Entre</i> | | | | | 0.05 |
| <i>Error</i> | 12 | | | | |
| <i>Total</i> | | | | | |

Tabla de efectos

| | |
|------------------------------------|---------------|
| α_1 Fármaco | 0 |
| α_2 Tratamiento psicológico | -4 |
| α_3 Grupo control | |
| | M = 10 |

CUESTIONARIO DEL SUPUESTO DE INVESTIGACIÓN 7: responder desde la pregunta 45 a la pregunta 53:

45. La metodología de la investigación se considera:

- A. Experimental porque se manipula la variable independiente y hay asignación aleatoria de las unidades experimentales a los grupos de la variable independiente.
- B. Experimental porque se manipula la variable independiente y hay asignación aleatoria en el orden de la administración de los grupos de la variable independiente.
- C. Cuasi-experimental porque aunque existe asignación aleatoria a los grupos de tratamiento, sin embargo no se manipula la variable independiente de tratamiento.
- D. No se puede responder a la pregunta.

46. La validez externa del estudio anterior queda garantizada ya que:

- A. Aleatoriamente los sujetos son asignados a las condiciones experimentales.
- B. Aleatoriamente son seleccionados los niveles de la variable independiente de tratamiento.
- C. Aleatoriamente los sujetos son seleccionados del conjunto de la población.
- D. No se puede responder a la pregunta.

47. La variable dependiente (o variables dependientes) es:

- A. Número de alteraciones paroxísticas.
- B. Número de trastornos que presenta el niño.
- C. La edad del niño.
- D. d) No se puede responder a la pregunta.

48. La variable independiente (o variables independientes) es:

- A. La cantidad de alteraciones paroxísticas que presenta el niño (ninguna, entre 3 y 4, más de cuatro).
- B. El grupo de intervención al que es sometido el niño (control, tratamiento psicológico, fármaco).
- C. La edad del niño (cinco años, más de cinco, menos de cinco).
- D. No se puede responder a la pregunta.

49. La Suma de Cuadrados del Tratamiento es:

- A. 96.
- B. 116.
- C. 506.
- D. No se puede responder a la pregunta.

50. En esta primera fase de su investigación la decisión estadística permite concluir que:

- A. Existen diferencias estadísticamente significativas ya que $F(2, 8) = 38$, $p < 0.05$.
- B. Existen diferencias estadísticamente significativas ya que $F(2, 6) = 43.8$, $p < 0.05$.
- C. Existen diferencias estadísticamente significativas ya que $F(2, 6) = 24$, $p < 0.05$.
- D. No se puede responder a la pregunta.

51. El investigador decide calcular una prueba de comparación de medias que le permita analizar el número total posible de comparaciones simples entre pares de medias. Cuántas comparaciones simples son posibles analizar sin ser redundantes:

- A. 3.
- B. 4.
- C. 2.
- D. d) No se puede responder a la pregunta.

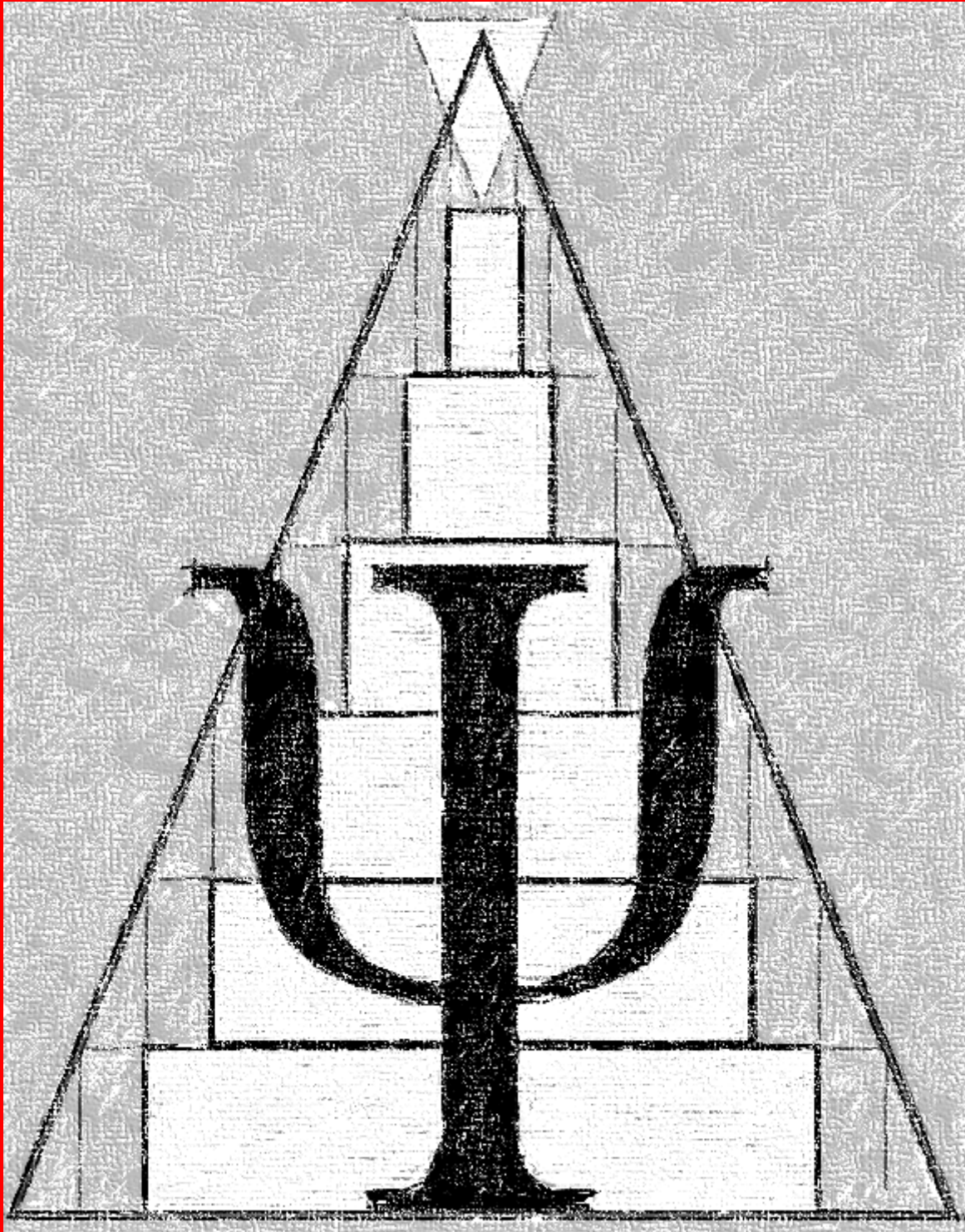
52. Si el investigador decide realizar comparaciones de medias, la prueba estadística más adecuada estará guiada por:

- A. Controlar la tasa de error de Tipo I y ser la más potente.
- B. Controlar la tasa de exceso y ser exacta.
- C. Controlar la tasa de error de Tipo I y trabajar con el menor Error de Tipo II.
- D. No se puede responder a la pregunta.

53. Si el investigador hubiese planteado la comparación de la puntuación media del grupo control frente a las puntuaciones medias de todos los demás grupos de tratamiento, realizando comparaciones simples que hubiesen estado definidas a priori, la prueba de comparación de medidas más adecuada sería:

- A. La prueba de Dunnett siempre que se realicen 2 comparaciones de medias.
- B. La prueba de Dunnett siempre que se realicen 3 comparaciones de medias.
- C. La prueba de Tukey siempre que se realicen más de 4 comparaciones de medias.
- D. No se puede responder a la pregunta.

Frías-Navarro, D. y Pascual-Soler, M. (Eds.), *Diseño de la investigación, análisis y redacción de los resultados*. Universidad de Valencia. España. DOI: <https://doi.org/10.17605/OSF.IO/KNGTP> Disponible en: osf.io/kngtp



**Edición 1^a, septiembre de 2020
DOI 10.17605/OSF.IO/KNGTP**

