Investigación sobre la investigación

Ciencia de la ciencia, metaciencia y sus métodos

Pablo Sánchez-Núñez
Universidad Carlos III de Madrid

26 de noviembre de 2026

Índice

Presentación 1
Usar R en este libro 4
Capítulo 1. Qué es la ciencia de la ciencia 6
Capítulo 2. Medir la ciencia 12
Capítulo 3. La crisis de reproducibilidad 19
Capítulo 4. Prácticas cuestionables y sesgos 25
Capítulo 5. Ciencia abierta y reformas 31
Capítulo 6. Evaluar la investigación 37
Capítulo 7. Fuentes de datos e infraestructuras 44
Capítulo 8. Bibliometría e indicadores 50
Capítulo 9. Redes y mapas de la ciencia 56
Capítulo 10. Minería de textos, modelos de temas y modelos de lenguaje 61
Capítulo 11. Síntesis de la evidencia y detección de sesgos 66
Capítulo 12. Diseños de investigación sobre la investigación 71
Laboratorio interactivo 77
Guía para el profesorado 79
Glosario 80
Referencias 84
Bibliografía 84

Presentación

ℹ Ficha del libro

- **Autor:** Pablo Sánchez-Núñez
- **ORCID:** https://orcid.org/0000-0001-7845-9506
- **Web del autor:** https://pablosancheznunez.com
- **Afiliación:** Universidad Carlos III de Madrid
- **DOI:** https://doi.org/10.5281/zenodo.24528120
- **Licencia:** CC BY 4.0
- **Versión web interactiva:** https://pablosancheznunez.com/ciencia-de-la-ciencia/
- **Cómo citarlo:** Sánchez-Núñez, P. (2026). *Investigación sobre la investigación: ciencia de la ciencia, metaciencia y sus métodos*. Universidad Carlos III de Madrid. Zenodo. https://doi.org/10.5281/zenodo.24528120

ℹ Sobre esta versión en PDF

Esta es la versión para leer e imprimir. El código de R aparece sin ejecutar, las preguntas de autoevaluación se muestran sin corrección automática y las actividades interactivas aparecen en papel, con su solución; los laboratorios solo funcionan en la web. Para ejecutar el código, hacer los ejercicios con corrección y responder a las preguntas, usa la versión web: https://pablosancheznunez.com/ciencia-de-la-ciencia/.

La ciencia es la forma más fiable que hemos inventado para conocer el mundo, y también es una actividad humana: tiene incentivos, sesgos, modas y fallos. La **ciencia de la ciencia** (*science of science*) estudia con métodos científicos cómo se produce, se comunica y se evalúa el conocimiento. La **investigación sobre la investigación** (*research on research*, RoR) y la **metaciencia** ponen el foco en mejorarla: detectar qué falla y comprobar si las reformas funcionan.

El libro tiene dos partes. La **primera** presenta los fundamentos: qué es el campo, cómo se mide la ciencia, la crisis de reproducibilidad, las prácticas cuestionables, las reformas y la evaluación. La **segunda** reúne las **técnicas y metodologías** con las que se investiga la investigación: fuentes de datos, bibliometría e indicadores, redes y mapas, minería de textos y modelos de lenguaje, revisiones sistemáticas y metaanálisis, detección de sesgos y diseños causales, replicaciones, análisis multiverso y métodos cualitativos y mixtos, con ejemplos resueltos sobre datos reales de OpenAlex.

Este libro está pensado para doctorandos y personal investigador de cualquier disciplina. Al terminarlo sabrás leer críticamente un indicador bibliométrico, estimar cuánto puedes fiarte de un resultado publicado, reconocer las prácticas cuestionables (incluidas las tuyas) y elegir las prácticas abiertas que hacen tu trabajo más creíble.

ℹ Cómo citar este libro

Sánchez-Núñez, P. (2026). *Investigación sobre la investigación: ciencia de la ciencia, metaciencia y sus métodos*. Universidad Carlos III de Madrid. Zenodo. https://doi.org/10.5281/zenodo.24528120

Autor: Pablo Sánchez-Núñez, Universidad Carlos III de Madrid
ORCID: 0000-0001-7845-9506
Web: pablosancheznunez.com

DOI: 10.5281/zenodo.24528120. Texto bajo licencia CC BY 4.0.

Mapa del libro

Capítulo	Pregunta
Parte I. Fundamentos	
1. Qué es la ciencia de la ciencia	¿Cómo se estudia científicamente la propia ciencia?
2. Medir la ciencia	¿Qué miden (y qué no) las citas y los indicadores?
3. La crisis de reproducibilidad	¿Cuánto podemos fiarnos de un resultado publicado?
4. Prácticas cuestionables y sesgos	¿Cómo se fabrican falsos positivos sin hacer trampa a propósito?
5. Ciencia abierta y reformas	¿Qué cambios mejoran la credibilidad y qué pruebas hay?
6. Evaluar la investigación	¿Cómo evaluar de forma responsable personas, proyectos e instituciones?

Capítulo	Pregunta
Parte II. Técnicas y metodologías	
7. Fuentes de datos e infraestructuras	¿De dónde salen los datos y qué sesgos tienen?
8. Bibliometría e indicadores	¿Cómo contar y normalizar producción e impacto con responsabilidad?
9. Redes y mapas de la ciencia	¿Cómo se relacionan documentos, personas y temas?
10. Minería de textos y modelos de lenguaje	¿Cómo analizar miles de textos científicos y validar el resultado?
11. Síntesis de la evidencia y sesgos	¿Cómo revisar, metaanalizar y detectar sesgo de publicación y errores?
12. Diseños de investigación	¿Cómo estimar efectos causales y medir la incertidumbre analítica?

Cómo está organizado cada capítulo

Cada capítulo empieza con sus **objetivos**, el **glosario** del capítulo y una **lectura obligatoria** con DOI. Después vienen el contenido, **laboratorios** interactivos, **actividades** de corrección inmediata, un **reto evaluable** sin solución en el libro y una **autoevaluación**. Los ejercicios de R se ejecutan en tu navegador: si nunca has usado R, empieza por Usar R en este libro.

Tu progreso se guarda solo en tu navegador. El libro funciona sin conexión una vez visitado, salvo los laboratorios que consultan datos en vivo.

Usar R en este libro

Varios capítulos tienen ejercicios cortos de R que se ejecutan **en tu navegador**, sin instalar nada. Los usarás para calcular indicadores y para simular cómo se generan resultados falsos. Este capítulo te explica lo justo en diez minutos; si ya usas R, puedes saltarlo.

i Objetivos

- Ejecutar y modificar código de R en el navegador.
- Guardar valores y textos en objetos y trabajar con vectores.

Cómo funcionan las celdas

Cada bloque de código tiene un botón **Run Code**. Puedes cambiar el código y volver a ejecutarlo; **Start Over** recupera la versión original. La primera vez que abres una página, el navegador descarga el motor de R (unos segundos).

```
citas <- c(45, 12, 8, 3, 0)  # citas de cinco artículos
sum(citas)                   # total
mean(citas)                  # media
median(citas)                # mediana: menos sensible a un artículo muy citado
```

Con `<-` se guarda un valor en un objeto. Un **vector** reúne varios valores con `c()`. Los textos van entre comillas:

```
bases <- c("Web of Science", "Scopus", "OpenAlex", "Dimensions")
length(bases)                    # cuántas hay
"OpenAlex" %in% bases            # ¿está OpenAlex?
sort(citas, decreasing = TRUE)   # citas de mayor a menor
```

Ejercicio 1

De 480 estudios replicados, 24 obtienen un resultado coherente con el original. Calcula la proporción de replicaciones exitosas.

i Ejercicio: completa los huecos _____

```
pertinentes <- 24
resultados <- ______
pertinentes / resultados
```

i Pista

Sustituye el hueco por el número total de estudios replicados.

Solución

```
pertinentes <- 24
resultados <- 480
pertinentes / resultados
```

0,05: una de cada veinte. Es un ejemplo inventado; en el capítulo 3 verás cifras reales.

Ejercicio 2

Comprueba si «Scopus» está en el vector `bases`.

Ejercicio: completa los huecos _____

```
"_______" %in% bases
```

Pista

Escribe el nombre exactamente igual, con mayúscula inicial.

Solución

```
"Scopus" %in% bases
```

`TRUE`. Si escribes «scopus» en minúsculas, R dice `FALSE`: distingue mayúsculas y minúsculas.

Autoevaluación

Responde en el papel y comprueba después con las explicaciones del recuadro. En la versión web estas preguntas se corrigen automáticamente. No cuentan para la nota.

1. En R, los textos se escriben entre comillas. *[V / F]*
2. Para crear un vector con varios valores se usa la función *[length() / c() / toupper()]*
3. `length(c("a", "b", "c"))` devuelve _____
4. Para hacer los ejercicios del libro hay que instalar R en el ordenador. *[V / F]*

Explicaciones

- Sin comillas, R buscaría un objeto con ese nombre.
- `c()` combina valores en un vector.
- Tres elementos.
- Se ejecutan en el navegador con webR.

Capítulo 1. Qué es la ciencia de la ciencia

Objetivos

- Distinguir ciencia de la ciencia, investigación sobre la investigación y metaciencia.
- Conocer los hitos y las preguntas clásicas del campo.
- Identificar las fuentes de datos y los métodos con que se estudia la ciencia.

Glosario del capítulo

Ciencia de la ciencia · Metaciencia · Investigación sobre la investigación · Cienciometría · Efecto Mateo · Ley de Price

Descargas

Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

Fortunato, S.; Bergstrom, C. T.; Börner, K.; Evans, J. A.; Helbing, D.; Milojević, S. et al. (2018). Science of science. *Science*, 359(6379), eaao0185.

Lecturas complementarias

- Price (1965) — El nacimiento del análisis de redes de citas.
- Merton (1968) — Por qué el reconocimiento se acumula en quien ya lo tiene.
- Ioannidis et al. (2015)
- Sinatra et al. (2016)
- Wuchty et al. (2007)

1. Tres nombres para un campo

Tres etiquetas que se solapan, con acentos distintos:

Nombre	Pregunta central	Tradición
Ciencia de la ciencia (*science of science*, SciSci)	¿Cómo funciona la ciencia como sistema: quién produce, colabora, cita e innova?	Cienciometría, sociología de la ciencia, física de redes
Investigación sobre la investigación (*research on research*, RoR)	¿Qué funciona para que la investigación sea mejor y más útil, y cómo deben decidir quienes la financian?	Política científica, evaluación
Metaciencia o metainvestigación	¿Son fiables los métodos y los resultados? ¿Qué reformas los mejoran?	Estadística, psicología, biomedicina

Lo que comparten es el giro **empírico**: no se limitan a opinar sobre cómo debería ser la ciencia, sino que reúnen datos (publicaciones, citas, financiación, replicaciones, encuestas) y los analizan como cualquier otro objeto de estudio. Fortunato et al. (2018) resumen el programa: entender los mecanismos que impulsan la ciencia para acelerar el descubrimiento y orientar las políticas científicas con evidencia.

2. Una breve historia

- **1939-1965.** John D. Bernal estudia la función social de la ciencia; Derek de Solla Price muestra que la literatura científica crece de forma exponencial y propone la **ley de Price**: una pequeña parte de los autores produce una gran parte de los trabajos. En 1965 dibuja la primera red de citas.
- **1955-1963.** Eugene Garfield propone el índice de citas y crea el *Science Citation Index*, la base de Web of Science.
- **1968.** Robert K. Merton describe el **efecto Mateo**: los investigadores ya reconocidos reciben más reconocimiento por contribuciones similares.
- **Décadas de 1970-1990.** Nace la **cienciometría** como disciplina, con revistas propias (*Scientometrics*, 1978).
- **2005.** Ioannidis publica *Why most published research findings are false*, punto de partida de la metaciencia contemporánea.
- **Desde 2010.** Datos masivos (Web of Science, Scopus, OpenAlex, Dimensions), crisis de reproducibilidad, ciencia abierta y una nueva comunidad interdisciplinar: el Research on Research Institute (RoRI, 2019) o los congresos de metaciencia.

3. Preguntas que se hace el campo

Ámbito	Ejemplos de preguntas	Un hallazgo
Colaboración	¿Los equipos producen mejor ciencia que las personas solas?	Los equipos dominan cada vez más la producción y reciben más citas (Wuchty et al., 2007).
Carreras	¿Cuándo hacen los investigadores su trabajo más influyente?	El trabajo más citado puede aparecer en cualquier punto de la carrera (Sinatra et al., 2016).
Impacto	¿Se puede predecir qué artículos serán muy citados?	Las citas siguen patrones regulares, pero con mucha incertidumbre individual (Wang et al., 2013).
Desigualdad	¿Se reparte el reconocimiento con justicia?	El efecto Mateo y las brechas de género en autoría y citas están bien documentados.
Fiabilidad	¿Cuántos resultados publicados se replican?	Bastante menos de lo que se suponía en varias disciplinas (capítulo 3).
Reforma	¿Funcionan el preregistro o los informes registrados?	Hay pruebas prometedoras, y se siguen evaluando (capítulo 5).

4. Datos y métodos

La ciencia deja un rastro digital enorme. Las **fuentes** habituales son las bases de datos bibliográficas (Web of Science, Scopus, Dimensions y la abierta OpenAlex), los registros de identificadores (Crossref, DataCite, ORCID, ROR), las bases de financiación, los registros de ensayos y de preregistros, las patentes y las encuestas a investigadores.

Los **métodos** combinan:

- **Análisis de redes**: citas, coautorías, palabras clave.
- **Modelos estadísticos y econométricos**: efectos de la financiación, de las políticas o de la movilidad.
- **Experimentos y cuasiexperimentos**: por ejemplo, asignar al azar revisores o formatos de revisión.
- **Replicaciones a gran escala**: repetir decenas de estudios para estimar su fiabilidad.
- **Métodos cualitativos**: entrevistas y etnografías de laboratorio, que explican lo que los datos no ven.

! Una advertencia de partida

Los datos sobre la ciencia también tienen sesgos: cobertura desigual por idiomas, regiones y disciplinas, errores en la atribución de autores y afiliaciones, y diferencias entre bases de datos. Antes de concluir algo sobre «la ciencia», pregúntate qué ciencia está en tus datos.

Caso práctico resuelto

i Convertir una inquietud en una pregunta de investigación sobre la investigación

Inquietud: «Me parece que en mi campo casi nadie comparte los datos».

1. **Pregunta descriptiva:** ¿qué proporción de artículos de mi campo publicados en 2024 incluye una declaración de disponibilidad de datos, y cuántos enlazan a un repositorio?
2. **Datos:** una muestra aleatoria de 200 artículos de las revistas principales (lista y criterios fijados antes).
3. **Codificación propia:** un libro de códigos con categorías (sin declaración, «a petición», repositorio enlazado, datos realmente accesibles).
4. **Análisis:** proporciones con intervalos de confianza; comparación por revista y por política editorial.
5. **Pregunta causal posterior:** ¿aumentó la disponibilidad después de que una revista adoptara una política obligatoria? (diferencias en diferencias, capítulo 12).

Ejercicios graduados

Básico. Convierte en pregunta investigable: «La revisión por pares no funciona».

Explicaciones

Por ejemplo: «¿Qué proporción de errores introducidos deliberadamente en un manuscrito detectan los revisores?» o «¿Coinciden los revisores de un mismo manuscrito en su recomendación?». Una pregunta investigable define qué se mide.

Avanzado. Formula una pregunta descriptiva y otra causal sobre una práctica de tu campo y esboza los datos que necesitarías.

Explicaciones

La descriptiva debe definir población, variable y fuente; la causal, la intervención, la comparación y un diseño posible (capítulo 12).

Actividades

Actividad: ¿Qué enfoque es?

Categorías: Ciencia de la ciencia · Investigación sobre la investigación · Metaciencia

- Modelar cómo se propaga una idea nueva por la red de coautorías
- Comprobar con un experimento si la revisión anónima reduce sesgos al financiar proyectos
- Replicar 100 estudios de psicología para estimar cuántos se confirman
- Estudiar si los equipos grandes producen ideas más disruptivas que los pequeños
- Evaluar si los informes registrados reducen el sesgo de publicación
- Analizar qué criterios usan las agencias para repartir fondos

Solución

- Modelar cómo se propaga una idea nueva por la red de coautorías → **Ciencia de la ciencia**
- Comprobar con un experimento si la revisión anónima reduce sesgos al financiar proyectos → **Investigación sobre la investigación**
- Replicar 100 estudios de psicología para estimar cuántos se confirman → **Metaciencia**
- Estudiar si los equipos grandes producen ideas más disruptivas que los pequeños → **Ciencia de la ciencia**
- Evaluar si los informes registrados reducen el sesgo de publicación → **Metaciencia**
- Analizar qué criterios usan las agencias para repartir fondos → **Investigación sobre la investigación**

Actividad: Ordena los hitos

- Bernal publica *The social function of science* (1939)
- Garfield propone el índice de citas (1955)
- Ioannidis publica *Why most published research findings are false* (2005)
- Merton describe el efecto Mateo (1968)
- Price dibuja la primera red de citas (1965)
- Se crea el Research on Research Institute (2019)

Solución

1. Bernal publica *The social function of science* (1939)
2. Garfield propone el índice de citas (1955)
3. Price dibuja la primera red de citas (1965)
4. Merton describe el efecto Mateo (1968)
5. Ioannidis publica *Why most published research findings are false* (2005)
6. Se crea el Research on Research Institute (2019)

Actividad: ¿Qué fuente usarías?

Categorías: Base bibliográfica · Registro de identificadores · Registro de ensayos o preregistros · Encuesta o entrevista

- Contar los artículos de una universidad entre 2015 y 2024
- Desambiguar a dos investigadoras con el mismo nombre
- Comprobar si un ensayo clínico publicó todos sus resultados previstos
- Saber por qué el personal investigador no comparte sus datos

Solución

- Contar los artículos de una universidad entre 2015 y 2024 → **Base bibliográfica**
- Desambiguar a dos investigadoras con el mismo nombre → **Registro de identificadores**. ORCID.
- Comprobar si un ensayo clínico publicó todos sus resultados previstos → **Registro de ensayos o preregistros**
- Saber por qué el personal investigador no comparte sus datos → **Encuesta o entrevista**

Práctica

Elige un artículo de tu campo y búscalo en OpenAlex. Anota qué datos ofrece (autores, afiliaciones, conceptos, citas, financiación) y cuáles faltan o son incorrectos. Esa es la materia prima, y la primera limitación, de cualquier estudio de ciencia de la ciencia.

Reto evaluable

Reto del capítulo 1 · Una pregunta de metaciencia sobre tu campo

1. Formula una **pregunta** de ciencia de la ciencia o de metaciencia sobre tu disciplina (por ejemplo: «¿Ha crecido el tamaño de los equipos en mi campo?»).
2. Indica qué **datos** necesitarías y de qué fuentes; justifica sus sesgos de cobertura.
3. Propón un **método** para responderla y describe un resultado que confirmaría tu hipótesis y otro que la refutaría.
4. Busca dos estudios que hayan abordado preguntas parecidas y resume qué encontraron.

Entrega: un informe hecho con la plantilla de entrega. Este reto **no tiene solución en el libro**: cada persona elige su pregunta y la corrección la hace el profesorado.

Autoevaluación

Responde en el papel y comprueba después con las explicaciones del recuadro. En la versión web estas preguntas se corrigen automáticamente. No cuentan para la nota.

1. La metaciencia se centra sobre todo en la fiabilidad de los métodos y los resultados, y en cómo mejorarlos. *[V / F]*
2. El fenómeno por el que los investigadores ya reconocidos reciben más reconocimiento se llama efecto _____
3. ¿Quién creó el *Science Citation Index*? *[Robert K. Merton / Eugene Garfield / Derek de Solla Price]*
4. Los datos bibliográficos cubren por igual todas las lenguas y regiones. *[V / F]*
5. Según Wuchty et al. (2007), la producción científica está cada vez más dominada por *[investigadores individuales / equipos / grandes empresas]*

6. La base de datos bibliográfica abierta y gratuita del listado es *[Web of Science / Scopus / OpenAlex]*
7. ¿Qué tienen en común los tres nombres del campo?

(a) Que solo usan métodos cualitativos (b) Que estudian la ciencia de forma empírica, con datos y métodos científicos (c) Que se limitan a la biomedicina

8. Replicar decenas de estudios para estimar cuántos se confirman es un método de la metaciencia. *[V / F]*

Explicaciones

- Fiabilidad y reforma son su núcleo.
- Por el evangelio: «al que tiene, se le dará».
- Garfield, en 1963 (propuesta en 1955).
- La cobertura favorece al inglés, a Norteamérica y Europa y a algunas disciplinas.
- Los equipos producen más y reciben más citas.
- OpenAlex es abierta y gratuita.
- El giro empírico es lo que los une.
- Las replicaciones a gran escala son centrales en metaciencia.

Recursos complementarios

Recurso	Para qué sirve
OpenAlex	Índice abierto de la literatura científica
Research on Research Institute	Red internacional de investigación sobre la investigación
Center for Open Science	Organización de referencia en metaciencia y ciencia abierta
Leiden Madtrics	Blog del CWTS (Universidad de Leiden) sobre indicadores y evaluación

Capítulo 2. Medir la ciencia

Objetivos

- Entender qué representan las citas y cómo se distribuyen.
- Calcular e interpretar indicadores como el índice h y los indicadores normalizados.
- Aplicar los principios del Manifiesto de Leiden para usar las métricas con responsabilidad.

Glosario del capítulo

Cita · Índice h · Factor de impacto · Normalización por campo · Distribución asimétrica · Altmetría · Manifiesto de Leiden

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

Hicks, D.; Wouters, P.; Waltman, L.; de Rijcke, S.; Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431.

Lecturas complementarias

- Hirsch (2005)
- Wang et al. (2013)
- Declaración de San Francisco sobre la Evaluación de la Investigación (DORA), 2012.

1. Qué es una cita

Una **cita** es una referencia de un documento a otro. Merton la veía como el pago de una deuda intelectual; otros autores la ven como un instrumento de persuasión. En la práctica se cita por muchos motivos: reconocer una idea, apoyar un argumento, describir un método, discrepar o, simplemente, porque todo el mundo cita ese trabajo. Por eso las citas miden **uso o visibilidad dentro de la comunidad científica**, no calidad ni verdad.

Tres hechos que conviene tener siempre presentes:

1. **Las citas se distribuyen de forma muy asimétrica.** Unos pocos artículos acumulan muchas citas y la mayoría recibe pocas o ninguna. La **media** está dominada por los extremos; la **mediana** y los percentiles describen mejor lo típico.
2. **Las prácticas de citación varían por campo**: un artículo de biología celular recibe, de media, muchas más citas que uno de matemáticas o de historia. Comparar recuentos brutos entre campos es comparar peras con manzanas.
3. **Las citas tardan en llegar** y su ritmo depende del campo. Los indicadores de ventanas cortas (dos años) favorecen a las disciplinas rápidas.

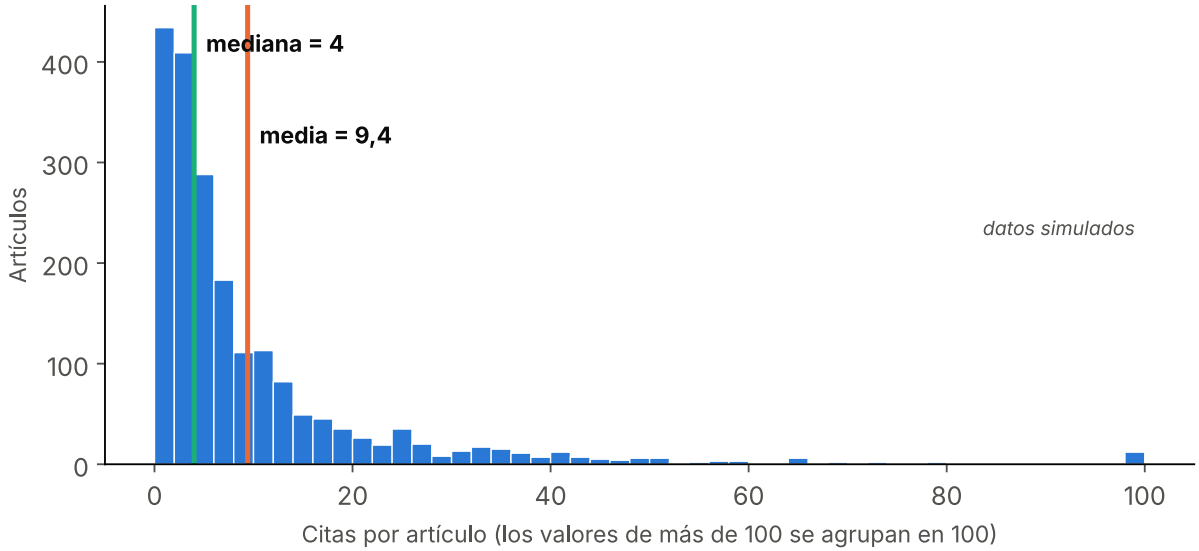

Figura 1: Distribución típica de las citas: muchos artículos con pocas citas y unos pocos con muchas. La media queda muy por encima de lo típico.

2. Indicadores más usados

Indicador	Qué mide	Problemas
Recuento de citas	Atención recibida por un trabajo	Depende del campo, la antigüedad y el tipo de documento
Factor de impacto (de la revista)	Citas medias recibidas por los artículos de una revista en dos años	Es de la **revista**, no del artículo; media de una distribución asimétrica; manipulable
Índice h (Hirsch, 2005)	El mayor h tal que h trabajos tienen al menos h citas	Crece con la edad académica; ignora las citas de los muy citados; no comparable entre campos
Citas normalizadas (p. ej. MNCS, FWCI)	Citas relativas a la media mundial del mismo campo, año y tipo	Dependen de cómo se definen los campos
Percentiles (p. ej. % en el 10 % más citado)	Posición en la distribución del campo	Necesitan muchos trabajos para ser estables
Altmetría	Menciones en redes, noticias, políticas, gestores	Mide atención, no calidad; fácil de inflar

El **índice h** es popular porque combina cantidad e impacto en un solo número fácil de calcular, pero precisamente por eso oculta mucho. Dos investigadoras con $h = 10$ pueden tener perfiles muy distintos.

3. Laboratorio: calcula un índice h

> **i** Laboratorio interactivo
>
> Escribe las citas de cada trabajo de un perfil (separadas por comas). El laboratorio las ordena, calcula el índice h y lo compara con la media y la mediana.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

4. Calcular con R

```
citas <- c(120, 45, 30, 22, 15, 12, 9, 8, 6, 5, 3, 2, 1, 0, 0)
ordenadas <- sort(citas, decreasing = TRUE)
h <- sum(ordenadas >= seq_along(ordenadas))
h
c(media = mean(citas), mediana = median(citas))
```

La línea clave compara cada recuento con su posición: `seq_along()` genera 1, 2, 3…, y el índice h es cuántas veces las citas igualan o superan la posición.

Ejercicio 2.1

Calcula el índice h de este perfil.

Ejercicio: completa los huecos _____

```
ord <- sort(perfil, decreasing = TRUE)
sum(ord >= _____(ord))
```

Pista

La función que genera la secuencia 1, 2, 3… de la longitud del vector es `seq_along()`.

Solución

```
ord <- sort(perfil, decreasing = TRUE)
sum(ord >= seq_along(ord))
```

h = 5: los cinco trabajos más citados tienen 25, 15, 9, 8 y 7 citas, todos al menos 5; el sexto tiene 4 citas, menos que 6.

Ejercicio 2.2

Un artículo de matemáticas tiene 12 citas; la media mundial de los artículos de matemáticas del mismo año y tipo es 4. Un artículo de biología celular tiene 30 citas y la media de su campo es 40. Calcula la **citación normalizada** (citas / media del campo) del artículo de matemáticas.

Ejercicio: completa los huecos _____

```
12 / _____
```

Pista

Divide entre la media de su campo.

Solución

```
12 / 4
```

3: el triple de lo esperado en su campo. El de biología, con 30 citas, obtiene 30 / 40 = 0,75, por debajo de lo esperado. Con recuentos brutos parecería al revés.

5. Usar las métricas con responsabilidad

El **Manifiesto de Leiden** (Hicks y others, 2015) propone diez principios. Los más citados:

1. La evaluación cuantitativa debe **apoyar** el juicio experto cualitativo, no sustituirlo.
2. Medir el desempeño en relación con la **misión** de la institución, el grupo o la persona.
3. Proteger la excelencia en la investigación **local** (incluida la publicada en otros idiomas).
4. Que los procesos de recogida y análisis de datos sean **abiertos, transparentes y sencillos**.
5. Permitir que las personas evaluadas **verifiquen** los datos y los análisis.
6. Tener en cuenta las **diferencias entre campos** en publicación y citación.
7. Basar la evaluación individual en el juicio cualitativo de la **trayectoria**.
8. Evitar la **falsa precisión** (por ejemplo, tres decimales en un factor de impacto).
9. Reconocer los **efectos sistémicos** de la evaluación y de los indicadores.
10. **Revisar y actualizar** los indicadores con regularidad.

La declaración **DORA** (2012) añade una recomendación concreta: no usar métricas de revista, como el factor de impacto, para evaluar artículos ni personas. El acuerdo europeo **CoARA** (2022) compromete a las instituciones firmantes a reformar sus evaluaciones en esa dirección (capítulo 6).

Ley de Goodhart

«Cuando una medida se convierte en objetivo, deja de ser una buena medida». Si se premia el número de artículos, aparecen el troceado de resultados y las revistas depredadoras; si se premian las citas, las autocitas y los círculos de citación.

Caso práctico resuelto

¿Qué indicador uso para esta decisión?

Una comisión debe repartir fondos internos entre tres departamentos (Historia, Química e Informática). Propuesta inicial: comparar el número total de citas. Problemas: las prácticas de citación difieren mucho por campo y el tamaño de los departamentos también.

Alternativa responsable:

1. Indicadores **normalizados** por campo (MNCS, PP(top 10 %)) y producción **por persona**.
2. Producción no indexada en las bases internacionales (libros en Historia, conferencias en Informática).
3. Informes cualitativos de cada departamento sobre su misión y sus resultados.
4. Un panel que use los indicadores como **apoyo**, no como fórmula.

Ejercicios graduados

Básico. ¿Por qué no es válido comparar el factor de impacto medio de las revistas de Historia y de Química?

Explicaciones

Porque las prácticas de publicación y citación son muy distintas entre campos, y el factor de impacto describe revistas, no la calidad de los trabajos de cada departamento.

Avanzado. Diseña un pequeño cuadro de indicadores para evaluar un grupo de investigación de tu área y justifica cada indicador con el Manifiesto de Leiden.

Explicaciones

Debe combinar indicadores normalizados, información cualitativa y la misión del grupo, explicar los límites de cada indicador y permitir que el grupo verifique los datos.

Actividades

Actividad: ¿Uso responsable o irresponsable del indicador?

Categorías: Responsable · Irresponsable

- Ordenar candidatos a una plaza por el factor de impacto medio de sus revistas
- Comparar el impacto de dos grupos de distintas disciplinas con citas normalizadas por campo, como complemento a una evaluación experta
- Comparar el índice h de una investigadora de 30 años con el de un catedrático de 65
- Publicar los datos y el código con que se calcularon los indicadores de una evaluación
- Informar de un factor de impacto con tres decimales para distinguir revistas
- Valorar también la publicación en español en revistas de relevancia local

Solución

- Ordenar candidatos a una plaza por el factor de impacto medio de sus revistas → **Irresponsable**. Métrica de revista aplicada a personas (DORA).
- Comparar el impacto de dos grupos de distintas disciplinas con citas normalizadas por campo, como complemento a una evaluación experta → **Responsable**
- Comparar el índice h de una investigadora de 30 años con el de un catedrático de 65 → **Irresponsable**. El índice h crece con la edad académica.
- Publicar los datos y el código con que se calcularon los indicadores de una evaluación → **Responsable**
- Informar de un factor de impacto con tres decimales para distinguir revistas → **Irresponsable**. Falsa precisión.
- Valorar también la publicación en español en revistas de relevancia local → **Responsable**

Actividad: ¿Qué indicador usarías?

Categorías: Mediana de citas · Citas normalizadas por campo · Percentil del 10 % más citado · Ninguno: juicio experto

- Resumir lo «típico» de un conjunto de artículos con unas pocas superestrellas
- Comparar el impacto de dos departamentos de física y de historia
- Saber qué proporción de la producción está entre lo más citado del mundo
- Decidir si un artículo de este año es valioso

Solución

- Resumir lo «típico» de un conjunto de artículos con unas pocas superestrellas → **Mediana de citas**
- Comparar el impacto de dos departamentos de física y de historia → **Citas normalizadas por campo**
- Saber qué proporción de la producción está entre lo más citado del mundo → **Percentil del 10 % más citado**
- Decidir si un artículo de este año es valioso → **Ninguno: juicio experto**. Aún no ha tenido tiempo de recibir citas.

Reto evaluable

Reto del capítulo 2 · Auditoría de un indicador

1. Elige un perfil (el tuyo, el de tu grupo o el de una revista de tu campo) y obtén sus datos en una fuente abierta (OpenAlex) y en otra (Google Académico, Scopus o Web of Science).
2. Calcula el recuento de citas, la mediana, el índice h y, si puedes, un indicador normalizado.
3. Compara las fuentes: ¿coinciden? ¿Por qué no?
4. Escribe una valoración del perfil aplicando **cinco** principios del Manifiesto de Leiden.

Entrega: un informe reproducible (con el código o la hoja de cálculo) hecho con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Las citas miden la calidad científica de un artículo. *[V / F]*
2. Si un perfil tiene citas 10, 8, 5, 4 y 3, su índice h es *[3 / 4 / 5 / 8]*
3. El factor de impacto es un indicador de la *[persona / revista / institución / disciplina]*
4. Para describir lo típico de una distribución de citas muy asimétrica, es mejor la _____
5. La declaración que pide no usar el factor de impacto para evaluar personas se llama *[Leiden / DORA / Helsinki / Budapest]*
6. Un artículo con 20 citas en un campo cuya media es 10 tiene una citación normalizada de _____
7. «Cuando una medida se convierte en objetivo, deja de ser una buena medida» es la ley de _____
8. ¿Por qué no se pueden comparar recuentos de citas entre campos?

(a) Porque las bases de datos solo cubren un campo (b) Porque las prácticas de citación y publicación varían mucho entre disciplinas (c) Porque en algunos campos está prohibido citar

Explicaciones

- Miden uso o visibilidad en la comunidad.
- Cuatro trabajos tienen al menos 4 citas; no hay cinco con 5 o más.
- Media de citas de la revista en dos años.
- La media se deja arrastrar por los muy citados.
- DORA, San Francisco, 2012.
- 20 / 10 = 2.
- Charles Goodhart, economista.
- Por eso existen los indicadores normalizados.

Recursos complementarios

Recurso	Para qué sirve
Manifiesto de Leiden (web)	Los diez principios y sus traducciones
DORA	Declaración y recursos para evaluar sin métricas de revista
OpenAlex	Datos abiertos de citas y autores
CWTS Leiden Ranking	Ejemplo de indicadores transparentes y normalizados

Capítulo 3. La crisis de reproducibilidad

Objetivos

- Distinguir reproducibilidad, replicabilidad y robustez.
- Conocer los grandes proyectos de replicación y sus resultados.
- Calcular cuánto podemos fiarnos de un resultado significativo según la potencia, el umbral y la probabilidad previa.

Glosario del capítulo

Reproducibilidad · Replicabilidad · Potencia estadística · Valor predictivo positivo · Falso positivo · Maldición del ganador

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

1. Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

Lecturas complementarias

- Ioannidis (2005)
- Button et al. (2013)

1. Tres palabras que no son sinónimos

Término	Pregunta	Se comprueba…
Reproducibilidad (computacional)	¿Obtengo los mismos resultados con **los mismos datos y el mismo código**?	Volviendo a ejecutar el análisis
Replicabilidad	¿Obtengo un resultado coherente con **datos nuevos** recogidos con el mismo método?	Repitiendo el estudio
Robustez	¿Se mantiene el resultado con **otros análisis razonables** de los mismos datos?	Variando decisiones analíticas

La terminología varía entre disciplinas (en algunas, «reproducibilidad» significa lo que aquí llamamos replicabilidad), así que conviene definirla siempre al escribir.

2. Qué han encontrado los proyectos de replicación

En 2015, la Open Science Collaboration repitió **100 estudios** publicados en tres revistas de psicología, con muestras más grandes y en contacto con los autores originales. El 97 % de los originales tenían resultados significativos; en las replicaciones, solo alrededor del **36 %**. Los tamaños del efecto de las replicaciones eran, de media, aproximadamente **la mitad** de los originales.

Proyectos posteriores, con métodos distintos, han encontrado tasas de replicación diversas en economía experimental, ciencias sociales publicadas en *Nature* y *Science*, o biología del cáncer preclínica. Las cifras no son comparables entre sí, y «replicar» puede definirse de varias formas (significación, tamaño del efecto,

intervalo), pero el mensaje común es claro: **un resultado publicado y significativo no es, por sí solo, una prueba sólida**.

> ℹ Una replicación fallida no es un fraude
>
> Un resultado puede no replicarse por azar en cualquiera de los dos estudios, por diferencias de población o de contexto, por errores, o porque el efecto original era un falso positivo o estaba inflado. Una sola replicación tampoco es concluyente: lo que importa es la evidencia acumulada.

3. Por qué tantos resultados no se replican

Ioannidis (2005) mostró con un modelo sencillo que la probabilidad de que un resultado significativo sea verdadero, el **valor predictivo positivo** (VPP), depende de tres cosas:

- La **probabilidad previa** de que la hipótesis sea cierta (expresada como razón R entre hipótesis verdaderas y falsas que se ponen a prueba).
- La **potencia** del estudio ($1 - \beta$): la probabilidad de detectar el efecto si existe.
- El **umbral de significación** α (normalmente 0,05).

$$\text{VPP} = (1 - \beta) \cdot R \,/\, [\,(1 - \beta) \cdot R + \alpha\,]$$

Con hipótesis arriesgadas (R bajo) y estudios con poca potencia, la mayoría de los «hallazgos» significativos pueden ser falsos aunque nadie haga trampas. Si además hay **sesgo** (prácticas que convierten resultados nulos en significativos, capítulo 4), el VPP baja todavía más.

Button et al. (2013) estimaron que la potencia mediana de los estudios de neurociencia rondaba el **21 %**. La poca potencia tiene un segundo efecto, la **maldición del ganador**: cuando un estudio pequeño sale significativo, lo hace casi siempre porque sobrestimó el efecto; por eso las replicaciones encuentran efectos más pequeños.

4. Laboratorio: ¿cuánto te fías de un resultado significativo?

> ℹ Laboratorio interactivo
>
> Imagina 1000 hipótesis puestas a prueba en un campo. Ajusta la proporción de hipótesis verdaderas, la potencia, el umbral y el sesgo, y mira cuántos resultados significativos son verdaderos.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

5. Calcular con R

```r
vpp <- function(R, potencia, alfa = 0.05) (potencia * R) / (potencia * R + alfa)
vpp(R = 0.1 / 0.9, potencia = 0.8)   # 10 % de hipótesis verdaderas, potencia alta
vpp(R = 0.1 / 0.9, potencia = 0.2)   # la misma probabilidad previa, potencia baja
```

Ejercicio 3.1

En un campo exploratorio, solo 1 de cada 20 hipótesis probadas es verdadera (R = 1/19). Con potencia del 80 % y α = 0,05, ¿cuál es el VPP?

> ℹ Ejercicio: completa los huecos _____

```r
vpp(R = 1 / 19, potencia = _____)
```

Pista

La potencia del 80 % se escribe 0.8.

Solución

```
vpp(R = 1 / 19, potencia = 0.8)
```

0,46: incluso con buena potencia, menos de la mitad de los resultados significativos serían verdaderos. En campos exploratorios, un único resultado significativo es una pista, no una conclusión.

Ejercicio 3.2: la maldición del ganador

Simulamos 2000 estudios pequeños (20 personas por grupo) de un efecto real pequeño (d = 0,2) y nos quedamos con los significativos. Completa el código para calcular el **tamaño del efecto medio** de los estudios significativos.

Ejercicio: completa los huecos _____

```
significativos <- estudios[estudios$p < 0.05, ]
nrow(significativos) / nrow(estudios)    # potencia observada
mean(significativos$_____)               # efecto medio publicado
```

Pista

La columna con el tamaño del efecto se llama `d`.

Solución

```
significativos <- estudios[estudios$p < 0.05, ]
nrow(significativos) / nrow(estudios)
mean(significativos$d)
```

Solo alrededor de uno de cada diez estudios sale significativo, y el efecto medio de esos estudios es varias veces mayor que el real (0,2): si solo se publican los significativos, la literatura exagera el efecto.

Caso práctico resuelto

Intentar reproducir un análisis publicado

Escoges un artículo con datos y código abiertos. Procedimiento:

1. **Descargar** datos y código del repositorio indicado.
2. **Reconstruir el entorno** (versiones de R o Python y paquetes, si se documentan).
3. **Ejecutar** el código sin modificarlo y anotar cada error.
4. **Comparar** cada cifra de las tablas y figuras principales con el artículo.
5. **Clasificar el resultado:** reproducible sin cambios, con cambios menores, con diferencias numéricas, no reproducible.
6. **Informar con cortesía** a los autores de las diferencias encontradas.

Resultado típico: las cifras se reproducen tras resolver dependencias no documentadas, lo que demuestra la importancia de fijar el entorno (renv, uv, contenedores).

Ejercicios graduados

Básico. ¿En qué se diferencian reproducibilidad y replicabilidad?

Explicaciones

Reproducibilidad: mismos datos y mismo análisis, mismos resultados. Replicabilidad: datos nuevos para responder la misma pregunta, resultados compatibles.

Avanzado. Intenta reproducir un análisis publicado de tu campo con datos abiertos y documenta el proceso con la clasificación del caso.

Explicaciones

Informa del tiempo empleado, de los obstáculos (datos incompletos, código ausente, versiones) y de las cifras que coinciden y las que no.

Actividades

Actividad: ¿Reproducibilidad, replicabilidad o robustez?

Categorías: Reproducibilidad · Replicabilidad · Robustez

- Vuelvo a ejecutar el código de un artículo con sus datos y obtengo las mismas tablas
- Repito el experimento con 300 personas nuevas
- Analizo los mismos datos excluyendo atípicos de otra forma razonable y el resultado se mantiene
- Un laboratorio de otro país repite el protocolo original
- No consigo obtener las cifras del artículo porque falta un paso del código

Solución

- Vuelvo a ejecutar el código de un artículo con sus datos y obtengo las mismas tablas → **Reproducibilidad**
- Repito el experimento con 300 personas nuevas → **Replicabilidad**
- Analizo los mismos datos excluyendo atípicos de otra forma razonable y el resultado se mantiene → **Robustez**
- Un laboratorio de otro país repite el protocolo original → **Replicabilidad**
- No consigo obtener las cifras del artículo porque falta un paso del código → **Reproducibilidad**

Actividad: ¿Qué hace más probable que un resultado sea un falso positivo?

Categorías: Lo hace más probable · Lo hace menos probable

- Muestras pequeñas con poca potencia
- Hipótesis muy sorprendentes, con poca probabilidad previa
- Un plan de análisis preregistrado
- Muchas decisiones analíticas tomadas después de ver los datos
- Un umbral de significación más estricto (0,005)
- Un efecto replicado de forma independiente

Solución

- Muestras pequeñas con poca potencia → **Lo hace más probable**
- Hipótesis muy sorprendentes, con poca probabilidad previa → **Lo hace más probable**
- Un plan de análisis preregistrado → **Lo hace menos probable**
- Muchas decisiones analíticas tomadas después de ver los datos → **Lo hace más probable**
- Un umbral de significación más estricto (0,005) → **Lo hace menos probable**
- Un efecto replicado de forma independiente → **Lo hace menos probable**

Reto evaluable

Reto del capítulo 3 · Intenta reproducir un artículo

1. Elige un artículo de tu campo que comparta **datos y código**.
2. Intenta **reproducir** una tabla o una figura principal. Documenta cada paso, cada obstáculo y el tiempo empleado.
3. Haz **un análisis de robustez**: cambia una decisión razonable (exclusiones, covariables, transformación) y comprueba si el resultado se mantiene.
4. Estima con el modelo de Ioannidis el VPP del hallazgo principal, justificando tus supuestos de probabilidad previa y potencia.

Entrega: informe y código con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Obtener los mismos resultados con los mismos datos y el mismo código es *[replicabilidad / reproducibilidad computacional / robustez]*
2. ¿Qué porcentaje aproximado de replicaciones fue significativo en el proyecto de la 1? *[10 % / 36 % / 70 % / 97 %]*
3. Una sola replicación fallida demuestra que el estudio original era fraudulento. *[V / F]*
4. Si la potencia baja, el valor predictivo positivo de un resultado significativo *[sube / baja / no cambia]*
5. Que los estudios pequeños significativos sobrestimen el efecto se conoce como la maldición del _____
6. Según Button et al. (2013), la potencia mediana en neurociencia rondaba el *[5 % / 21 % / 50 % / 80 %]*
7. El VPP depende de la probabilidad previa de que la hipótesis sea cierta. *[V / F]*
8. ¿Por qué los tamaños del efecto de las replicaciones suelen ser menores?

(a) Porque los equipos que replican son menos competentes (b) Porque los originales publicados, seleccionados por ser significativos, tienden a sobrestimar el efecto (c) Porque las replicaciones usan muestras más pequeñas

Explicaciones

- Mismos datos, mismo código.
- En torno al 36 %, frente al 97 % de los originales.
- Hay muchas causas posibles, incluido el azar.
- Con menos potencia, una mayor proporción de los significativos son falsos.
- *Winner's curse*.
- Alrededor del 21 %.
- Es la *R* del modelo de Ioannidis.
- Selección por significación + muestras pequeñas = efectos inflados.

Recursos complementarios

Recurso	Para qué sirve
OSF	Compartir datos, código y materiales para que otros reproduzcan tu trabajo
Redes nacionales de reproducibilidad (UKRN)	Directorio de redes nacionales, incluida la Red Española de Reproducibilidad (SPRN)
FORRT	Recursos docentes sobre reproducibilidad y ciencia abierta

Capítulo 4. Prácticas cuestionables y sesgos

Objetivos

- Reconocer las prácticas de investigación cuestionables y por qué parecen inocentes.
- Comprobar con simulaciones cómo los «grados de libertad del investigador» fabrican falsos positivos.
- Entender el sesgo de publicación y los incentivos que lo sostienen.

Glosario del capítulo

Prácticas cuestionables · *p-hacking* · HARKing · Grados de libertad del investigador · Sesgo de publicación · Parada opcional

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

Simmons, J. P.; Nelson, L. D.; Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.

Lecturas complementarias

- John et al. (2012)
- Fanelli (2010)
- Smaldino y McElreath (2016)

1. Entre la buena práctica y el fraude

La mala conducta científica en sentido estricto se resume en **FFP**: *fabrication* (inventar datos), *falsification* (manipularlos) y *plagiarism*. Es grave pero relativamente rara. Mucho más frecuentes son las **prácticas de investigación cuestionables** (*questionable research practices*, QRP): decisiones que no son fraude, que a menudo se toman de buena fe y que, sin embargo, aumentan la probabilidad de publicar resultados falsos.

Práctica	En qué consiste
Informar solo de las variables que salen	Medir varias variables dependientes y contar solo las significativas
Parada opcional	Mirar los datos a medida que llegan y dejar de recoger cuando $p < 0{,}05$
Excluir datos a conveniencia	Decidir qué observaciones son «atípicas» después de ver el resultado
Elegir el análisis que funciona	Probar varias covariables, transformaciones o pruebas y contar solo una
HARKing (*hypothesizing after the results are known*)	Presentar como hipótesis previa lo que se descubrió al analizar
Redondear *p*	Escribir $p = 0{,}05$ cuando es 0,054
No publicar los resultados nulos	Guardar en el cajón los estudios que «no salen»

John et al. (2012) encuestaron a más de 2000 psicólogos con incentivos para responder con sinceridad: muchas de estas prácticas eran admitidas por una proporción considerable, en algunos casos por la mayoría.

2. El jardín de los senderos que se bifurcan

Simmons et al. (2011) demostraron con simulaciones que combinar unas pocas decisiones flexibles, cada una defendible por sí sola, puede elevar la tasa de falsos positivos del 5 % nominal a **más del 60 %**. Gelman y Loken llamaron a este problema el «jardín de los senderos que se bifurcan»: aunque solo se haga un análisis, la elección depende de los datos, y otros datos habrían llevado a otro camino.

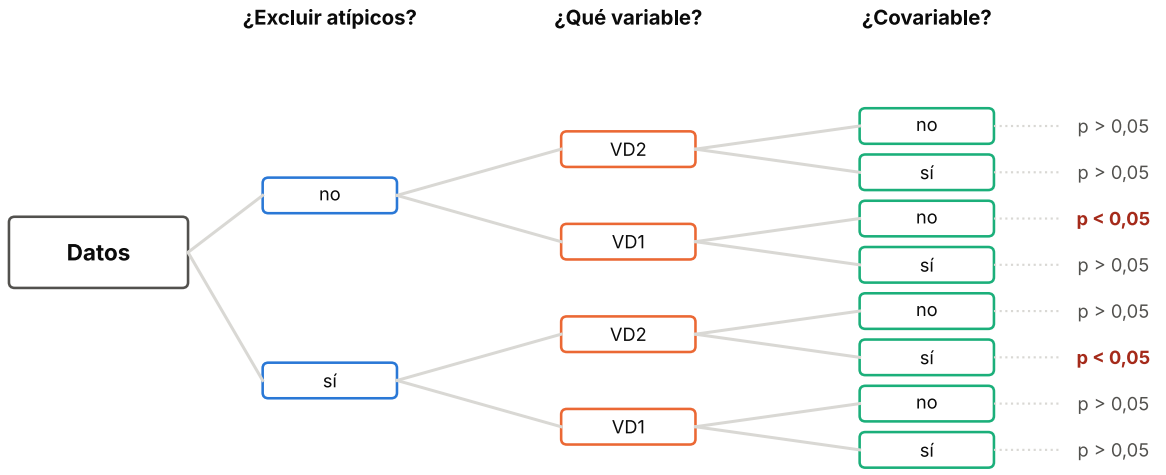

Figura 2: El jardín de los senderos que se bifurcan: tres decisiones binarias crean ocho análisis posibles con los mismos datos.

3. Laboratorio: fabrica un falso positivo sin hacer trampa

> ℹ Laboratorio interactivo
>
> Simulamos 2000 estudios en los que **no hay ningún efecto real** (dos grupos idénticos). Con un análisis honesto y único, alrededor del 5 % sale significativo por azar. Activa las decisiones flexibles y mira cómo sube la tasa de falsos positivos.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

4. Sesgo de publicación

El **sesgo de publicación** aparece cuando la probabilidad de publicar un estudio depende de su resultado. Fanelli (2010) encontró que la proporción de artículos que confirman su hipótesis es altísima en muchas disciplinas y mayor en las ciencias sociales que en las físicas. Si los resultados nulos se quedan en el cajón, la literatura publicada es una muestra sesgada de lo investigado: los metaanálisis heredan el sesgo y sobrestiman los efectos.

Algunas señales de alarma en una literatura:

- Demasiados resultados positivos para la potencia de los estudios.
- Valores p que se amontonan justo por debajo de 0,05.
- Estudios pequeños con efectos grandes y estudios grandes con efectos pequeños (asimetría en el **gráfico de embudo** de un metaanálisis).

5. Incentivos: la selección natural de la mala ciencia

¿Por qué persisten estas prácticas si casi nadie quiere engañar? Smaldino y McElreath (2016) mostraron con un modelo evolutivo que, si se premia sobre todo el **número de publicaciones positivas**, los laboratorios que usan métodos de menor potencia y más flexibles producen más «éxitos», reciben más recursos y forman a más investigadores que copian esos métodos. Sin cambiar los incentivos, la mala ciencia se selecciona sola, aunque las personas actúen de buena fe. Es el puente hacia los capítulos 5 y 6.

6. Calcular con R

Ejercicio 4.1: dos variables dependientes

Simulamos estudios sin efecto real en los que se miden dos variables independientes entre sí y se informa de la que salga. Completa el código para calcular la tasa de falsos positivos.

i Ejercicio: completa los huecos _____

```
sale <- p_vd1 < 0.05 | p_vd2 < ______
mean(sale)
```

i Pista

El umbral es el mismo para las dos variables: 0.05.

Solución

```
sale <- p_vd1 < 0.05 | p_vd2 < 0.05
mean(sale)
```

Alrededor del 10 %: el doble del 5 % nominal. Con variables independientes, la probabilidad es $1 - 0{,}95^2 \approx 0{,}0975$. Con más variables o más decisiones, sigue subiendo.

Caso práctico resuelto

i Detectar el jardín de los senderos que se bifurcan en un análisis propio

Un análisis de la relación entre sueño y rendimiento admite estas decisiones razonables:

- Excluir o no a quienes duermen menos de 4 horas (¿error de respuesta?).
- Medir el rendimiento con la nota media o con la nota del último semestre.
- Controlar o no por horas de trabajo remunerado.
- Usar el sueño continuo o en categorías.

Hay $2 \times 2 \times 2 \times 2 = 16$ análisis posibles. Si se prueban varios y se informa solo del significativo, la tasa de falsos positivos aumenta mucho. Remedios: **preregistrar** las decisiones, informar de todas (análisis multiverso, capítulo 12) o separar claramente lo **confirmatorio** de lo **exploratorio**.

Ejercicios graduados

Básico. ¿Qué práctica cuestionable es formular la hipótesis después de ver los resultados y presentarla como previa?

Explicaciones

HARKing (*hypothesizing after the results are known*).

Avanzado. Enumera las decisiones analíticas abiertas en tu proyecto actual y calcula cuántos análisis posibles generan.

Explicaciones

Multiplica las opciones de cada decisión, indica cuáles fijarías en un preregistro y cuáles explorarías, informando de ello.

Actividades

Actividad: ¿Mala conducta, práctica cuestionable o buena práctica?

Categorías: Mala conducta (FFP) · Práctica cuestionable · Buena práctica

- Inventar los datos de diez participantes que faltaban
- Dejar de recoger datos en cuanto el resultado es significativo
- Decidir y registrar antes de empezar el criterio de exclusión de atípicos
- Presentar como hipótesis inicial algo que se descubrió explorando los datos
- Retocar una imagen de un gel para que la banda se vea mejor
- Informar de todas las variables medidas, también las no significativas
- Copiar sin citar un párrafo de otro artículo

Solución

- Inventar los datos de diez participantes que faltaban → **Mala conducta (FFP)**
- Dejar de recoger datos en cuanto el resultado es significativo → **Práctica cuestionable**
- Decidir y registrar antes de empezar el criterio de exclusión de atípicos → **Buena práctica**
- Presentar como hipótesis inicial algo que se descubrió explorando los datos → **Práctica cuestionable**. HARKing.
- Retocar una imagen de un gel para que la banda se vea mejor → **Mala conducta (FFP)**
- Informar de todas las variables medidas, también las no significativas → **Buena práctica**
- Copiar sin citar un párrafo de otro artículo → **Mala conducta (FFP)**

ℹ Escenario: El análisis que no sale

Tu estudio preregistrado no encuentra el efecto esperado (p = 0,11). Tu directora sugiere probar otras opciones.

1. El resultado

¿Qué haces?

- Pruebo varias covariables hasta que sale significativo y lo publico así ⟶ paso 2
- Mantengo el análisis preregistrado como principal y añado análisis exploratorios etiquetados como tales ⟶ paso 3

2. Sendero bifurcado

(mal final)

Has convertido un resultado nulo en un probable falso positivo. Además, se desvía del preregistro sin decirlo: quien lea el artículo no podrá valorarlo.

3. Transparencia

Un análisis exploratorio con una covariable sale significativo. ¿Cómo lo presentas?

- En el resumen, como hallazgo principal ⟶ paso 4
- Como exploratorio, que habría que confirmar con datos nuevos, y publico también el resultado nulo principal ⟶ paso 5

4. Titular engañoso

(final mejorable)

Aunque lo etiquetes en el texto, destacarlo en el resumen da una impresión equivocada de la evidencia.

5. Resultado nulo publicado

(buen final)

Publicas un resultado nulo bien hecho, que reduce el sesgo de publicación, y una hipótesis nueva para el siguiente estudio. Los informes registrados (capítulo 5) protegen justo este caso.

Reto evaluable

! Reto del capítulo 4 · El jardín de tu propio análisis

1. Toma un análisis que hayas hecho (o uno publicado en tu campo) y **lista todas las decisiones** que se tomaron: exclusiones, transformaciones, covariables, pruebas, variables.
2. Para cada una, señala si se tomó **antes o después** de ver los datos.
3. Si tienes los datos, haz un **análisis multiverso** sencillo: repite el análisis con las combinaciones razonables y muestra cuántas dan el mismo resultado.
4. Propón cómo blindarías el próximo estudio frente a esas decisiones.

Entrega: informe y código con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Las prácticas de investigación cuestionables son siempre fraude intencionado. *[V / F]*
2. Presentar como hipótesis previa lo descubierto tras analizar los datos se llama _____

3. Según Simmons et al. (2011), combinar varias decisiones flexibles puede elevar la tasa de falsos positivos por encima del *[6 % / 15 % / 60 % / 95 %]*
4. Mirar los datos a medida que llegan y parar cuando $p < 0{,}05$ se llama parada _____
5. Con dos variables independientes y $\alpha = 0{,}05$, la probabilidad de al menos un falso positivo es aproximadamente *[5 % / 10 % / 25 %]*
6. El sesgo de publicación hace que los metaanálisis tiendan a *[subestimar los efectos / sobrestimar los efectos / no verse afectados]*
7. Smaldino y McElreath (2016) explican la persistencia de la mala ciencia sobre todo por *[la falta de inteligencia de los investigadores / los incentivos que premian publicar muchos resultados positivos / el coste de los ordenadores]*
8. ¿Qué tienen en común FFP?

(a) Son prácticas aceptables en algunas disciplinas (b) Son las tres formas de mala conducta en sentido estricto: fabricación, falsificación y plagio (c) Son tipos de sesgo de publicación

Explicaciones

- A menudo se toman de buena fe.
- *Hypothesizing After the Results are Known.*
- Más del 60 % en su combinación más extrema.
- *Optional stopping.*
- $1 - 0{,}95^2 \approx 0{,}0975$.
- Solo se ve la parte significativa (y exagerada).
- Selección natural de métodos por incentivos.
- *Fabrication, falsification, plagiarism.*

Recursos complementarios

Recurso	Para qué sirve
Retraction Watch	Seguimiento de retractaciones y casos de mala conducta
Código europeo de conducta para la integridad en la investigación (ALLEA)	Marco de referencia en Europa
statcheck	Comprueba la coherencia de los resultados estadísticos publicados

Capítulo 5. Ciencia abierta y reformas

Objetivos

- Conocer las principales reformas propuestas por la metaciencia y las pruebas sobre su eficacia.
- Distinguir preregistro, informe registrado y análisis exploratorio.
- Planificar un estudio con prácticas abiertas proporcionadas a su riesgo y su contexto.

Glosario del capítulo

Ciencia abierta · Preregistro · Informe registrado · Datos FAIR · Insignias de prácticas abiertas · Preprint

Descargas

Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

Munafò, M. R.; Nosek, B. A.; Bishop, D. V. M.; Button, K. S.; Chambers, C. D.; Percie du Sert, N. et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021.

Lecturas complementarias

- Nosek et al. (2018)
- Chambers (2013)
- Kaplan y Irvin (2015)
- Scheel et al. (2021)
- Kidwell y others (2016)

1. Un manifiesto en cinco frentes

Munafò et al. (2017) agrupan las medidas para una ciencia más reproducible en cinco ámbitos:

Ámbito	Medidas
Métodos	Proteger frente al sesgo cognitivo (enmascaramiento), mejorar la formación metodológica y estadística, colaborar en equipos grandes
Información	Preregistrar estudios, seguir guías de comunicación (CONSORT, PRISMA, ARRIVE…)
Reproducibilidad	Compartir datos, código y materiales; promover la transparencia
Evaluación	Diversificar la revisión por pares y apoyar la revisión posterior a la publicación
Incentivos	Premiar la apertura y la calidad, no solo el número de publicaciones positivas

2. Preregistro e informes registrados

Un **preregistro** es un plan del estudio (hipótesis, diseño, muestra, variables, análisis) registrado con fecha antes de ver los datos, en un repositorio público como OSF Registries o AsPredicted. Su función no es prohibir la exploración, sino **distinguir** lo confirmatorio de lo exploratorio (Nosek et al., 2018). Desviarse del plan es legítimo si se declara y se justifica.

Un **informe registrado** (*Registered Report*) va un paso más allá: la revista revisa la introducción, los métodos y el plan de análisis **antes** de recoger los datos y, si los acepta, se compromete a publicar el artículo **sea cual sea el resultado** (Chambers, 2013). Así desaparecen a la vez el incentivo para el *p-hacking* y el sesgo de publicación. Hoy los ofrecen cientos de revistas.

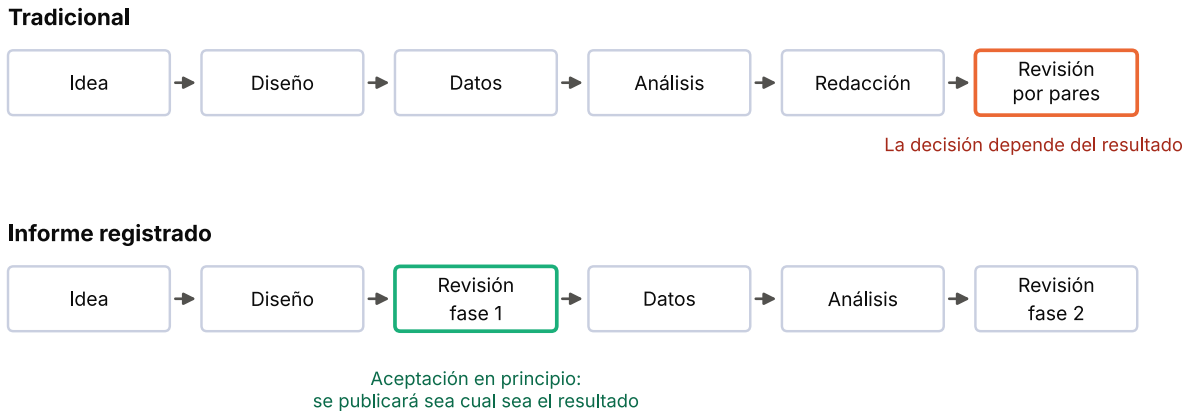

Figura 3: Del estudio tradicional al informe registrado: la revisión y la aceptación se adelantan a la recogida de datos.

3. Compartir datos, código y materiales

Compartir permite **reproducir**, **reutilizar** y **detectar errores**. Los principios **FAIR** piden que los datos sean localizables, accesibles, interoperables y reutilizables, lo que no siempre significa abiertos: «tan abiertos como sea posible, tan cerrados como sea necesario», por ejemplo con datos personales. Buenas prácticas mínimas:

- Depositar en un repositorio con **DOI** (Zenodo, OSF, un repositorio institucional o temático), no en una web personal.
- Incluir un **README** con la descripción de variables, la procedencia y la licencia.
- Compartir el **código** que va de los datos brutos a las figuras, con las versiones del software.
- Citar los datos y el software como cualquier otra fuente.

4. ¿Funcionan las reformas?

La metaciencia también se evalúa a sí misma. Algunas pruebas:

- Tras introducir **insignias** por compartir datos en *Psychological Science*, la proporción de artículos con datos abiertos aumentó mucho más que en revistas comparables (Kidwell y others, 2016).
- En los grandes ensayos clínicos financiados por el NHLBI estadounidense sobre fármacos o suplementos para enfermedades cardiovasculares, la proporción de ensayos con resultado positivo cayó notablemente tras exigirse el **registro previo** del resultado principal en 2000 (Kaplan e Irvin, 2015).
- Los **informes registrados** publican una proporción mucho mayor de resultados que no apoyan la hipótesis que los artículos tradicionales (Scheel et al., 2021).

Son pruebas observacionales en su mayoría, y la investigación sobre qué intervenciones funcionan, en qué disciplinas y con qué costes, sigue abierta. La reforma también tiene riesgos: burocracia, «lavado de ciencia abierta» (*open washing*) o desigualdades entre quienes tienen recursos para cumplir y quienes no.

5. Laboratorio: ¿qué prácticas abiertas necesita tu estudio?

ℹ Escenario: Planifica tu estudio

Vas a empezar un estudio empírico de tu tesis.

1. Tipo de estudio

¿Qué tipo de estudio es?

- Confirmatorio: pongo a prueba una hipótesis concreta ⟶ paso 2
- Exploratorio: busco patrones sin hipótesis previa ⟶ paso 5
- Análisis de datos secundarios que ya existen ⟶ paso 6

2. Confirmatorio

Tienes una hipótesis y un diseño. ¿Hay una revista de tu campo que acepte informes registrados?

- Sí ⟶ paso 3
- No o no lo sé ⟶ paso 4

3. Informe registrado

(buen final)

Envía la fase 1 (introducción, métodos, plan de análisis y cálculo de potencia) antes de recoger datos. Comparte datos y código al publicar.

4. Preregistro

(buen final)

Preregistra hipótesis, muestra (con su justificación), variables y análisis en OSF o AsPredicted. Declara en el artículo cualquier desviación.

5. Exploratorio

(buen final)

No hace falta un preregistro confirmatorio, pero sí transparencia: preséntalo como exploratorio, comparte los datos y el código, y considera reservar una parte de los datos para confirmar lo encontrado.

6. Datos secundarios

¿Has explorado ya los datos?

- No todavía ⟶ paso 7
- Sí, ya los conozco ⟶ paso 8

7. Preregistro de datos secundarios

(buen final)

Existen plantillas de preregistro para datos secundarios: registra el plan antes de analizar y declara qué sabías de los datos.

8. Ya los conoces

(final mejorable)

Un preregistro tendría menos valor, pero aún puedes declarar qué análisis decidiste antes y cuáles después, compartir el código y hacer análisis de robustez o multiverso.

Caso práctico resuelto

Preparar un preregistro en OSF

Para un estudio sobre el efecto de un taller de escritura en la productividad de doctorandos:

1. **Hipótesis:** quienes hagan el taller enviarán más manuscritos en seis meses.
2. **Diseño:** asignación aleatoria entre los inscritos (taller ahora o en el semestre siguiente).
3. **Variables:** manuscritos enviados (resultado principal), satisfacción (secundario).
4. **Tamaño muestral** con análisis de potencia y regla de parada.
5. **Análisis principal** exacto (modelo y covariables) y criterios de exclusión.
6. **Análisis exploratorios** declarados como tales.

El preregistro se fecha y se hace público (o embargado) antes de recoger los datos. Desviarse es legítimo si se declara y se justifica.

Ejercicios graduados

Básico. ¿Prohíbe el preregistro hacer análisis no previstos?

Explicaciones

No: permite hacerlos, pero exige declararlos como exploratorios y distinguirlos de los confirmatorios.

Avanzado. Redacta un preregistro con la plantilla de OSF para un estudio que podrías hacer tú solo con datos existentes.

Explicaciones

Con datos existentes, hay que declarar qué se conoce ya de los datos y cómo se evita que ese conocimiento condicione el análisis (por ejemplo, separando una muestra de validación).

Actividades

Actividad: ¿A qué problema responde cada reforma?

Categorías: Sesgo de publicación · Flexibilidad analítica · Errores no detectados · Incentivos

- Informes registrados
- Preregistro del plan de análisis
- Compartir datos y código
- Valorar en las convocatorias los conjuntos de datos y el software publicados
- Revisión posterior a la publicación en plataformas abiertas

Solución

- Informes registrados → **Sesgo de publicación**. También frenan la flexibilidad analítica.
- Preregistro del plan de análisis → **Flexibilidad analítica**
- Compartir datos y código → **Errores no detectados**
- Valorar en las convocatorias los conjuntos de datos y el software publicados → **Incentivos**
- Revisión posterior a la publicación en plataformas abiertas → **Errores no detectados**

Actividad: Ordena las fases de un informe registrado

- Diseñar el estudio y el plan de análisis
- Enviar la fase 1 a la revista
- Enviar la fase 2 con los resultados
- Publicar, sea cual sea el resultado
- Recibir la aceptación en principio
- Recoger y analizar los datos

Solución

1. Diseñar el estudio y el plan de análisis
2. Enviar la fase 1 a la revista
3. Recibir la aceptación en principio
4. Recoger y analizar los datos
5. Enviar la fase 2 con los resultados
6. Publicar, sea cual sea el resultado

Reto evaluable

Reto del capítulo 5 · Tu preregistro

1. Escribe el **preregistro** completo del próximo estudio de tu tesis con la plantilla de OSF o de AsPredicted (hipótesis, diseño, justificación de la muestra, variables, análisis, criterios de exclusión).
2. Incluye un **plan de gestión de datos** de una página: dónde depositarás datos y código, con qué licencia y qué no podrás compartir y por qué.
3. Señala **dos riesgos** de tu plan (por ejemplo, un análisis que dependa de los datos) y cómo los gestionarás.

Entrega: el preregistro (sin necesidad de publicarlo) con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Preregistrar un estudio prohíbe hacer análisis exploratorios. *[V / F]*
2. En un informe registrado, la revista acepta el artículo en principio *[después de ver los resultados / antes de que se recojan los datos / solo si el resultado es significativo]*
3. Los datos FAIR siempre tienen que ser abiertos. *[V / F]*
4. La guía de comunicación para revisiones sistemáticas es *[CONSORT / PRISMA / ARRIVE]*

5. Según Kidwell y others (2016), las insignias por compartir datos *[no tuvieron efecto / se asociaron a un gran aumento de artículos con datos abiertos / redujeron la calidad de los datos]*
6. Un plan de estudio registrado con fecha antes de ver los datos es un _____
7. Desviarse del preregistro es legítimo si se declara y se justifica. *[V / F]*
8. ¿Qué dos problemas elimina a la vez un informe registrado?

(a) La falta de financiación y la burocracia (b) El incentivo para buscar significación y el sesgo de publicación (c) Los errores de código y el plagio

Explicaciones

- Solo pide distinguir lo exploratorio de lo confirmatorio.
- Aceptación en principio tras la fase 1.
- «Tan abiertos como sea posible, tan cerrados como sea necesario».
- PRISMA (CONSORT, ensayos; ARRIVE, animales).
- Aumento muy superior al de revistas comparables.
- *Preregistration.*
- La transparencia es lo esencial.
- La publicación no depende del resultado.

Recursos complementarios

Recurso	Para qué sirve
OSF Registries	Preregistrar estudios
Registered Reports (COS)	Revistas que aceptan informes registrados
EQUATOR Network	Guías de comunicación por tipo de estudio
Zenodo	Depositar datos y software con DOI

Capítulo 6. Evaluar la investigación

Objetivos

- Comprender cómo los sistemas de evaluación moldean lo que se investiga y cómo se publica.
- Conocer las iniciativas de reforma: DORA, el Manifiesto de Leiden, los Principios de Hong Kong y CoARA.
- Preparar y valorar evidencias de una trayectoria investigadora más allá de las métricas.

Glosario del capítulo

Evaluación responsable · CoARA · DORA · Principios de Hong Kong · CV narrativo · Revisión por pares · CRediT

Descargas

Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

Moher, D.; Bouter, L.; Kleinert, S.; Glasziou, P.; Sham, M. H.; Barbour, V. et al. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, 18(7), e3000737.

Lecturas complementarias

- Hicks y others (2015)
- Compromisos del Acuerdo para la reforma de la evaluación de la investigación (CoARA), 2022.
- Smaldino y McElreath (2016)

1. Lo que se evalúa es lo que se hace

La evaluación decide quién obtiene una plaza, un proyecto, una promoción o una acreditación. Por eso no solo **mide** la investigación: la **moldea**. Si la evaluación cuenta artículos en revistas de alto factor de impacto, el personal investigador tiende a trocear resultados, a perseguir temas de moda, a evitar los riesgos y a no dedicar tiempo a la docencia, los datos, el software, la revisión por pares o la transferencia. El capítulo 4 mostró que esos incentivos pueden seleccionar, sin mala intención, los métodos menos fiables.

2. El movimiento por una evaluación responsable

Iniciativa	Año	Idea central
DORA	2012	No usar métricas de revista (como el factor de impacto) para evaluar artículos ni personas; valorar todos los resultados de la investigación
Manifiesto de Leiden	2015	Diez principios para usar indicadores como apoyo, no como sustituto, del juicio experto
Principios de Hong Kong	2020	Evaluar a las personas por prácticas que favorecen la integridad

Iniciativa	Año	Idea central
CoARA	2022	Compromiso de instituciones europeas para reformar su evaluación, con plazos

Los **Principios de Hong Kong** (Moher et al., 2020) proponen valorar:

1. Las prácticas responsables en todo el ciclo de la investigación.
2. La comunicación completa y transparente, sea cual sea el resultado.
3. Las prácticas de ciencia abierta.
4. Una amplia gama de investigación, incluidas las replicaciones y las síntesis.
5. Otras contribuciones: revisión por pares, mentoría, divulgación, intercambio de conocimiento.

El **Acuerdo de CoARA** compromete a sus firmantes, entre otras cosas, a reconocer la diversidad de contribuciones y carreras, a basar la evaluación sobre todo en la revisión cualitativa con un uso responsable de los indicadores, a abandonar los usos inadecuados de las métricas de revista y del índice h, y a evitar los rankings de universidades en la evaluación de la investigación.

3. Herramientas para evaluar de otro modo

- **CV narrativos**: formatos con apartados en los que la persona explica sus contribuciones al conocimiento, a otras personas, a la comunidad y a la sociedad, con evidencias seleccionadas en lugar de listas exhaustivas.
- **Taxonomía CRediT**: 14 roles (conceptualización, curación de datos, análisis formal, software, supervisión…) para declarar quién hizo qué en un trabajo colectivo.
- **Evidencias diversas**: conjuntos de datos y software con DOI, protocolos, informes de revisión, materiales docentes abiertos, impacto en políticas.
- **Revisión por pares bien diseñada**: criterios explícitos, paneles diversos, formación de los evaluadores y, en algunos casos, sorteos parciales entre las propuestas que superan un umbral de calidad, que algunas agencias ya ensayan.

> **i** La revisión por pares también se estudia
>
> Sus problemas están documentados: baja concordancia entre revisores, sesgos, lentitud y dificultad para detectar errores. Experimentos como los de revisión abierta, doble anónima o con formación específica forman parte de la investigación sobre la investigación (capítulo 1).

4. Evaluar no es solo medir: un caso

ℹ Escenario: Formas parte de una comisión de selección

Una universidad firmante de CoARA convoca una plaza. Hay dos candidaturas finales.

1. Primera reunión

Un miembro propone ordenar las candidaturas por la suma de factores de impacto de sus publicaciones. ¿Qué respondes?

- Me parece objetivo y rápido: lo apoyo → paso 2
- Propongo valorar los criterios de la convocatoria con las evidencias aportadas → paso 3

2. Métrica de revista

(mal final)

El factor de impacto describe revistas, no personas ni artículos, y la propia institución se ha comprometido a no usarlo así (DORA y CoARA).

3. Criterios y evidencias

La candidata A tiene más artículos; la candidata B tiene menos, pero ha publicado un software muy usado, un conjunto de datos con DOI y un informe registrado con resultado nulo. ¿Cómo procedes?

- Leo las contribuciones seleccionadas de cada una y las valoro según el perfil de la plaza → paso 5
- Descarto el software y los datos porque no son publicaciones → paso 4

4. Contribuciones invisibles

(mal final)

Ignorar el software, los datos o los resultados nulos castiga justo las prácticas que hacen la ciencia más fiable.

5. Juicio experto informado

(buen final)

Has evaluado contribuciones, no contenedores. Los indicadores pueden informar, pero la decisión se basa en la lectura y en el perfil de la plaza.

Caso práctico resuelto

ℹ Redactar un CV narrativo a partir de un CV tradicional

Un investigador pasa de una lista de 40 publicaciones a un CV narrativo de las convocatorias que siguen los principios de CoARA:

1. **Contribución al conocimiento:** tres resultados destacados, explicando su importancia y su papel en cada uno (con CRediT), no solo dónde se publicaron.
2. **Desarrollo de otras personas:** dirección de tesis, mentoría, formación.
3. **Contribución a la comunidad investigadora:** revisión, edición, organización de congresos, ciencia abierta (datos y software publicados con DOI).
4. **Contribución a la sociedad:** transferencia, divulgación, asesoramiento.

Cada afirmación va acompañada de evidencias verificables (DOI, enlaces) y, si se usan indicadores, contextualizados por campo.

Ejercicios graduados

Básico. ¿Qué diferencia principal hay entre un CV narrativo y uno tradicional?

Explicaciones

El narrativo explica la contribución y su importancia con evidencias en varias dimensiones; el tradicional enumera resultados, sobre todo publicaciones.

Avanzado. Convierte tu CV en un CV narrativo de dos páginas con las cuatro secciones del caso.

Explicaciones

Cada sección debe incluir ejemplos concretos y verificables, explicar tu papel y evitar métricas sin contexto (como el factor de impacto de las revistas).

Actividades

Actividad: ¿Qué Principio de Hong Kong valora esta evidencia?

Categorías: Prácticas responsables · Comunicación completa · Ciencia abierta · Investigación diversa · Otras contribuciones

- Un protocolo con plan de gestión de datos y aprobación ética
- Publicar un resultado nulo con todos los análisis
- Depositar datos y código en Zenodo con DOI
- Una replicación independiente de un estudio influyente
- Dirigir tesis y revisar artículos para revistas

Solución

- Un protocolo con plan de gestión de datos y aprobación ética → **Prácticas responsables**
- Publicar un resultado nulo con todos los análisis → **Comunicación completa**
- Depositar datos y código en Zenodo con DOI → **Ciencia abierta**
- Una replicación independiente de un estudio influyente → **Investigación diversa**
- Dirigir tesis y revisar artículos para revistas → **Otras contribuciones**

Actividad: ¿Qué rol CRediT es?

Categorías: Conceptualización · Curación de datos · Análisis formal · Software · Supervisión · Redacción: revisión y edición

- Formuló la pregunta y los objetivos del estudio
- Limpió, documentó y mantuvo los datos
- Aplicó los modelos estadísticos
- Programó el paquete que hace los cálculos
- Dirigió el trabajo del equipo
- Revisó críticamente el manuscrito

Solución

- Formuló la pregunta y los objetivos del estudio → **Conceptualización**
- Limpió, documentó y mantuvo los datos → **Curación de datos**
- Aplicó los modelos estadísticos → **Análisis formal**
- Programó el paquete que hace los cálculos → **Software**
- Dirigió el trabajo del equipo → **Supervisión**
- Revisó críticamente el manuscrito → **Redacción: revisión y edición**

Actividad: ¿Compatible con CoARA?

Categorías: Compatible · Incompatible

- Valorar la mentoría y la docencia en la carrera investigadora
- Exigir un índice h mínimo para concurrir a una plaza
- Usar la posición de la universidad en un ranking internacional para repartir fondos internos
- Pedir un CV narrativo con cinco contribuciones seleccionadas y explicadas

Solución

- Valorar la mentoría y la docencia en la carrera investigadora → **Compatible**
- Exigir un índice h mínimo para concurrir a una plaza → **Incompatible**
- Usar la posición de la universidad en un ranking internacional para repartir fondos internos → **Incompatible**
- Pedir un CV narrativo con cinco contribuciones seleccionadas y explicadas → **Compatible**

Reto evaluable

! Reto del capítulo 6 · Tu trayectoria en formato narrativo

1. Redacta un **CV narrativo** de dos páginas con cuatro apartados: contribuciones al conocimiento, a la formación de otras personas, a la comunidad investigadora y a la sociedad.
2. Selecciona **cinco contribuciones** (no solo artículos) y explica para cada una qué aportó y tu papel con roles **CRediT**.
3. Analiza la convocatoria de evaluación más relevante para ti (una ayuda, una acreditación, una plaza): ¿qué criterios aplica? ¿Son coherentes con DORA y CoARA?
4. Propón **tres cambios** concretos a esa convocatoria y justifícalos con la literatura del libro.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. La evaluación de la investigación solo mide; no influye en cómo se investiga. *[V / F]*
2. La iniciativa europea de 2022 que compromete a las instituciones a reformar su evaluación se llama _____
3. Los Principios de Hong Kong se centran en evaluar a las personas investigadoras para fomentar la *[productividad / integridad / internacionalización]*
4. La taxonomía que describe quién hizo qué en un trabajo se llama *[ORCID / CRediT / DORA]*
5. Según CoARA, deben evitarse los rankings de universidades en la evaluación de la investigación. *[V / F]*
6. Un CV narrativo sustituye la lista exhaustiva de publicaciones por contribuciones seleccionadas y explicadas. *[V / F]*
7. Un sorteo parcial entre propuestas que superan un umbral de calidad es *[una práctica prohibida / una técnica que algunas agencias ensayan para repartir fondos / lo mismo que evaluar por métricas]*
8. ¿Por qué valorar los resultados nulos y las replicaciones?

(a) Porque son más fáciles de publicar (b) Porque son necesarios para una ciencia fiable y la evaluación tradicional los desincentiva (c) Porque tienen más citas

Explicaciones

- Lo que se evalúa es lo que se hace.
- *Coalition for Advancing Research Assessment.*
- *Fostering research integrity.*
- *Contributor Roles Taxonomy.*
- Es uno de sus compromisos.
- Es su rasgo central.
- Varias agencias han probado sorteos parciales.
- Corrigen el sesgo de publicación.

Recursos complementarios

Recurso	Para qué sirve
CoARA	Coalición europea para la reforma de la evaluación
DORA	Declaración y herramientas para evaluar sin factor de impacto
CRediT	Los 14 roles de contribución

Recurso	Para qué sirve
Résumé for Research and Innovation (UKRI)	Un modelo de CV narrativo

Capítulo 7. Fuentes de datos e infraestructuras

Objetivos

- Conocer las principales fuentes de datos para investigar la investigación: bases bibliográficas, identificadores persistentes, registros y repositorios.
- Comparar su cobertura, sus licencias y sus sesgos, y elegir la adecuada para cada pregunta.
- Leer y explorar datos reales de OpenAlex con R.

Glosario del capítulo

Base de datos bibliográfica · Cobertura · Identificador persistente · Desambiguación de autores · Metadatos abiertos

Descargas

Plantilla de entrega del reto (.qmd) Datos de OpenAlex del capítulo (.json)

Lectura obligatoria

Visser, M.; van Eck, N. J.; Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20-41.

1. Un mapa de las fuentes

La investigación sobre la investigación se apoya en datos que producen otros: editoriales, agencias de registro de DOI, bases de datos, financiadores, repositorios. Ninguna fuente es neutral ni completa; conocer qué recoge cada una es el primer paso metodológico.

Tipo	Ejemplos	Para qué	Acceso
Bases bibliográficas multidisciplinares	Web of Science, Scopus, Dimensions, OpenAlex	Producción, citas, colaboración, temas	Comerciales (las tres primeras); abierta CC0 (OpenAlex)
Registros de metadatos de DOI	Crossref, DataCite	Metadatos de artículos, datos y software; referencias abiertas	Abiertos
Identificadores persistentes	ORCID (personas), ROR (instituciones), DOI (objetos), Crossref Funder Registry	Desambiguar y enlazar	Abiertos
Bases temáticas	PubMed/MEDLINE, Europe PMC, arXiv, RePEc, ERIC	Campos concretos, texto completo en algunos casos	Abiertas en su mayoría
Integridad y correcciones	Retraction Watch (integrada en Crossref desde 2023), PubPeer	Retractaciones, comentarios posteriores	Abiertos
Ciencia abierta	OSF, registros de ensayos (ClinicalTrials.gov, EU CTR), Zenodo	Preregistros, datos, materiales	Abiertos

Tipo	Ejemplos	Para qué	Acceso
Atención en línea	Altmetric, Overton (documentos de políticas)	Menciones en medios, redes y políticas	Comerciales
Financiación y personal	Bases de proyectos (CORDIS, AEI), encuestas de carrera	Insumos y resultados	Variable

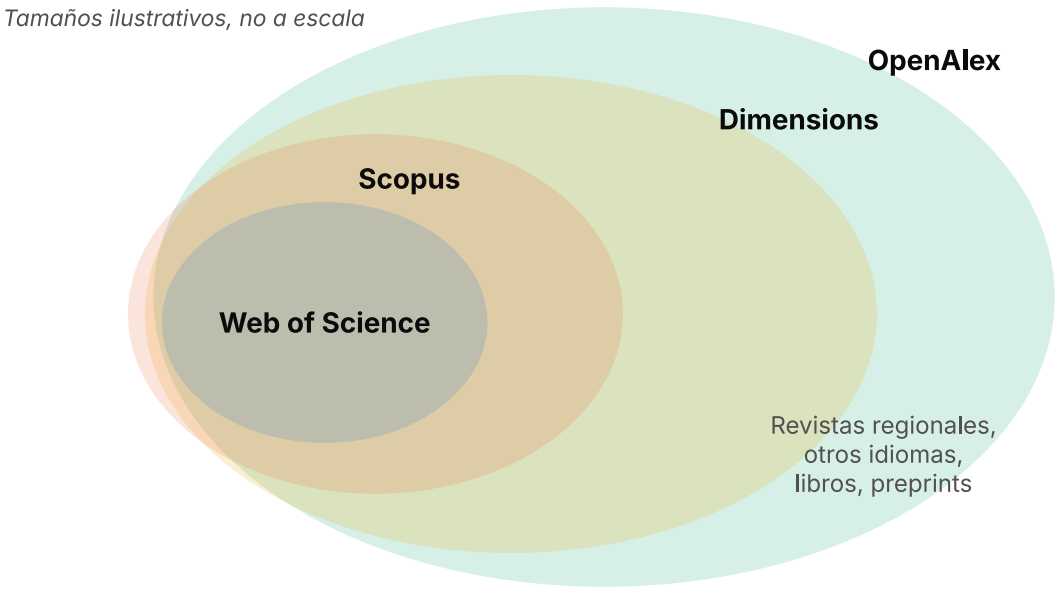

Figura 4: Cada fuente cubre una parte distinta del universo de publicaciones: ninguna lo contiene todo, y las diferencias no son aleatorias.

2. Cobertura y sesgos

Visser, van Eck y Waltman (2021) compararon a gran escala cinco fuentes. Sus conclusiones principales: las fuentes con criterios de selección (Web of Science, Scopus) son más pequeñas y selectivas; las de indexación amplia (Dimensions, Crossref, la extinta Microsoft Academic, hoy sustituida en gran medida por OpenAlex) cubren muchos más documentos, pero con metadatos de calidad más variable, por ejemplo en afiliaciones o tipos de documento. Las diferencias de cobertura afectan a los resultados: un ranking o un indicador puede cambiar según la fuente.

Sesgos que conviene comprobar siempre:

- **Idioma y región**: las fuentes selectivas infrarrepresentan la investigación publicada en lenguas distintas del inglés y en revistas regionales.
- **Disciplina**: las humanidades y parte de las ciencias sociales publican en libros y en revistas nacionales poco indexadas.
- **Tipo de documento**: artículos, revisiones, editoriales y cartas se clasifican de forma distinta en cada fuente.
- **Tiempo**: la cobertura retrospectiva es desigual y los metadatos se actualizan continuamente.
- **Desambiguación**: asignar trabajos a personas e instituciones es un problema algorítmico con errores; ORCID y ROR lo reducen, no lo eliminan.

3. Buenas prácticas con las fuentes

1. Justifica la elección de la fuente según la pregunta (cobertura temática, años, tipo de documento).
2. **Documenta la consulta exacta** y la fecha de descarga; guarda los datos brutos.

3. Usa **identificadores persistentes** para unir fuentes (DOI, ORCID, ROR) en lugar de títulos o nombres.
4. Valida una muestra a mano (por ejemplo, 100 registros) para estimar errores de clasificación.
5. Si es posible, **repite el análisis con una segunda fuente** como prueba de robustez.
6. Prefiere fuentes con **metadatos abiertos**: permiten que otros reproduzcan el estudio (principios de la *Barcelona Declaration on Open Research Information*, 2024).

Ejemplo resuelto: explorar una muestra de OpenAlex

El archivo del capítulo contiene los 80 artículos más citados que citan al gran proyecto de replicación en psicología de la 1. Así se explora:

```
d <- jsonlite::fromJSON("datos/metaciencia_openalex.json")$trabajos
nrow(d)
table(d$campo)                      # campo científico principal (clasificación de
OpenAlex)
summary(d$citas)                    # distribución de citas: muy asimétrica
n_autores <- sapply(d$autores, nrow)
c(media = mean(n_autores), mediana = median(n_autores))
```

Ejercicios

Ejercicio 7.1

¿En qué **año** se publicaron más artículos de la muestra?

Ejercicio: completa los huecos _____

```
tab <- table(d$anio)
names(tab)[which._____(tab)]
```

Pista

`which.max()` devuelve la posición del valor máximo.

Solución

```
tab <- table(d$anio)
names(tab)[which.max(tab)]
```

Ejercicio 7.2

¿Qué proporción de artículos tiene **al menos un autor de un país distinto de Estados Unidos** (código `US`)? Redondea a dos decimales.

Ejercicio: completa los huecos _____

```
otro_pais <- sapply(d, function(w) {
  paises <- unlist(lapply(w$autores, function(a) a$paises))
  any(paises != "_____")
})
round(mean(otro_pais), 2)
```

Pista

El código ISO de Estados Unidos es `US`.

Solución

```
otro_pais <- sapply(d, function(w) {
  paises <- unlist(lapply(w$autores, function(a) a$paises))
  any(paises != "US")
})
round(mean(otro_pais), 2)
```

Ojo: los autores sin país registrado no cuentan, lo que ilustra cómo los metadatos incompletos afectan a los indicadores de internacionalización.

Actividades

Actividad: ¿Qué fuente usarías?

Categorías: OpenAlex o Crossref · ORCID o ROR · Retraction Watch · Registros de ensayos u OSF

- Analizar con datos abiertos y reproducibles las citas de 50 000 artículos
- Desambiguar a qué institución pertenece cada autor
- Estudiar cuántos artículos retractados siguen citándose
- Comparar los resultados publicados con los planificados en preregistros

Solución

- Analizar con datos abiertos y reproducibles las citas de 50 000 artículos → **OpenAlex o Crossref**
- Desambiguar a qué institución pertenece cada autor → **ORCID o ROR**
- Estudiar cuántos artículos retractados siguen citándose → **Retraction Watch**
- Comparar los resultados publicados con los planificados en preregistros → **Registros de ensayos u OSF**

Actividad: ¿Qué sesgo de cobertura afecta?

Categorías: Idioma y región · Disciplina · Tipo de documento · Desambiguación

- Un estudio sobre producción en humanidades que solo cuenta artículos de revista
- Un ranking de países basado en una fuente que indexa pocas revistas en español
- Dos «J. García» distintos fusionados en un único perfil
- Editoriales contadas como artículos de investigación

Solución

- Un estudio sobre producción en humanidades que solo cuenta artículos de revista → **Disciplina**
- Un ranking de países basado en una fuente que indexa pocas revistas en español → **Idioma y región**
- Dos «J. García» distintos fusionados en un único perfil → **Desambiguación**
- Editoriales contadas como artículos de investigación → **Tipo de documento**

Reto evaluable

Reto del capítulo 7 · Dos fuentes, una pregunta

1. Formula una pregunta sencilla sobre tu campo (por ejemplo, evolución de la producción en un tema desde 2015).
2. Respóndela con dos fuentes (por ejemplo, OpenAlex y la base a la que tenga acceso tu universidad). Documenta las consultas y la fecha.
3. Compara los resultados, explica las diferencias por cobertura, clasificación o desambiguación, y valida a mano una muestra de 50 registros.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Según Visser et al. (2021), las fuentes selectivas como Web of Science cubren más documentos que Dimensions. *[V / F]*
2. El identificador persistente para instituciones de investigación es *[ORCID / ROR / DOI]*
3. Los datos de OpenAlex se publican con licencia _____
4. Un indicador puede cambiar según la fuente de datos utilizada. *[V / F]*
5. Asignar correctamente los trabajos a cada persona es el problema de la _____
6. La base de datos de retractaciones integrada en Crossref es *[PubPeer / Retraction Watch / Overton]*
7. Conviene documentar la consulta exacta y la fecha de descarga. *[V / F]*
8. ¿Por qué conviene repetir un análisis con una segunda fuente?

(a) Para tener más citas (b) Para comprobar que las conclusiones no dependen de la cobertura de una sola fuente (c) Porque lo exige la ley

Explicaciones

- Al revés: las de indexación amplia cubren más.
- *Research Organization Registry.*
- Dominio público.
- Por eso se justifica la fuente.
- *Author name disambiguation.*
- Desde 2023.
- Reproducibilidad.
- Prueba de robustez.

Capítulo 8. Bibliometría e indicadores

Objetivos

- Calcular e interpretar indicadores de producción y de impacto con recuento completo y fraccionario.
- Normalizar las citas por campo, año y tipo de documento, y usar indicadores de percentiles.
- Aplicar los principios del Manifiesto de Leiden al uso responsable de los indicadores.

Glosario del capítulo

Recuento completo · Recuento fraccionario · MNCS · Indicador de percentil · Ventana de citación · Manifiesto de Leiden

Descargas

Plantilla de entrega del reto (.qmd) Datos de OpenAlex del capítulo (.json)

Lectura obligatoria

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365-391.

Lectura complementaria

- Hicks y others (2015)

1. Contar publicaciones: completo o fraccionario

Cuando un artículo tiene autores de varias instituciones, ¿a quién se atribuye? Hay dos opciones principales:

- **Recuento completo**: cada institución (o país, o persona) suma 1. La suma de todos los créditos supera el número de artículos.
- **Recuento fraccionario**: el artículo se reparte; con 4 autores, cada uno recibe 1/4. La suma total coincide con el número de artículos.

Artículo	Autores (institución)	Completo: A / B	Fraccionario: A / B
1	A, A, B	1 / 1	0,67 / 0,33
2	A	1 / 0	1 / 0
3	B, B, B, A	1 / 1	0,25 / 0,75
Total		**3 / 2**	**1,92 / 1,08**

El recuento completo premia la colaboración amplia; el fraccionario es más adecuado para comparar productividad entre unidades con prácticas de colaboración distintas (Waltman, 2016).

2. Medir el impacto de citación

Las citas dependen del **campo** (en biomedicina se cita mucho más que en matemáticas), del **año** (los trabajos antiguos han tenido más tiempo) y del **tipo de documento** (las revisiones se citan más). Por eso los indicadores de impacto comparables se **normalizan**:

- **Citación normalizada de un trabajo**: sus citas divididas entre la media de citas de los trabajos del mismo campo, año y tipo. Un valor de 1 equivale a la media mundial.
- **MNCS** (*mean normalized citation score*): la media de las citaciones normalizadas de un conjunto de trabajos.

- **Indicadores de percentiles**: la proporción de trabajos de una unidad que están en el 10 % (o el 1 %) más citado de su campo y año, **PP(top 10 %)**. Son más robustos que las medias, porque una sola publicación muy citada no los dispara.
- **Ventana de citación**: contar las citas en un periodo fijo (por ejemplo, tres años tras la publicación) para comparar cohortes.

Waltman (2016) repasa además los problemas de los indicadores populares: el **factor de impacto** describe revistas, no artículos, y el **índice h** mezcla producción e impacto, depende de la edad académica y del campo, y no se puede normalizar bien.

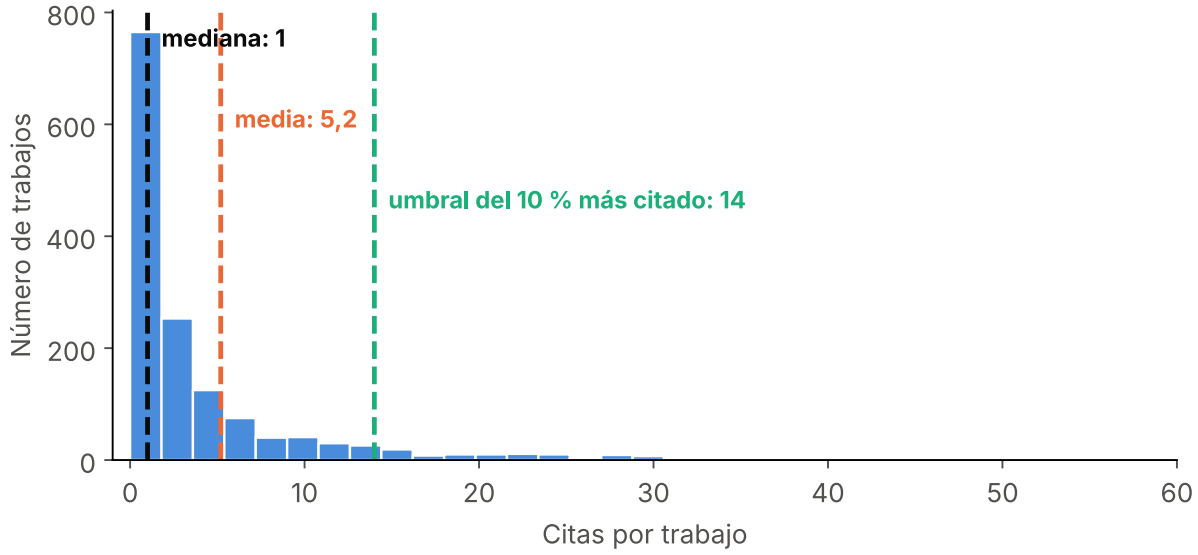

Figura 5: Las citas siguen una distribución muy asimétrica: la media queda muy por encima de la mediana, por eso los percentiles describen mejor el impacto que las medias.

3. Uso responsable: el Manifiesto de Leiden

Hicks y others (2015) propusieron diez principios para evaluar con indicadores. Entre ellos: la evaluación cuantitativa debe **apoyar** el juicio experto, no sustituirlo; hay que medir el desempeño en relación con la **misión** de la unidad; proteger la investigación **relevante localmente**; mantener los datos y el análisis **abiertos y transparentes**; permitir que los evaluados **verifiquen** los datos; tener en cuenta las **diferencias entre campos**; evaluar a las personas con un juicio cualitativo de su trayectoria; evitar la **falsa precisión**; reconocer los **efectos sistémicos** de los indicadores (incentivos y adaptación); y **revisar** los indicadores periódicamente.

Ejemplo resuelto: recuento fraccionario por país

```
d <- jsonlite::fromJSON("datos/metaciencia_openalex.json", simplifyVector =
FALSE)$trabajos
credito <- list()
for (w in d) {
  autores <- Filter(function(a) length(a$paises) > 0, w$autores)
  for (a in autores) for (p in a$paises)          # cada autor reparte su parte entre
sus países
    credito[[p]] <- c(credito[[p]], 1 / length(autores) / length(a$paises))
}
frac <- sort(sapply(credito, sum), decreasing = TRUE)
completo <- sort(table(unlist(lapply(d, function(w) unique(unlist(lapply(w$autores,
`[[`, "paises")))))), decreasing = TRUE)
```

```r
head(round(frac, 1), 6)
head(completo, 6)
```

Fíjate en que la suma del recuento fraccionario es, como máximo, el número de artículos con país conocido, mientras que el completo cuenta el mismo artículo varias veces.

Laboratorio: medias, medianas y percentiles

i Laboratorio interactivo

Dos grupos publican 20 artículos cada uno. El grupo B tiene un único artículo excepcionalmente citado. Mueve sus citas y observa qué indicador cambia.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

Ejercicios

Ejercicio 8.1 · Recuento fraccionario

Un artículo tiene 5 autores: 3 de la institución A y 2 de la B. ¿Qué crédito fraccionario recibe **A**?

i Ejercicio: completa los huecos _____

```r
autores <- c("A", "A", "A", "B", "B")
sum(autores == "A") / ______(autores)
```

i Pista

Divide entre el número total de autores: `length(autores)`.

Solución

```r
autores <- c("A", "A", "A", "B", "B")
sum(autores == "A") / length(autores)
```

0,6. Con recuento completo, A y B recibirían 1 cada una.

Ejercicio 8.2 · MNCS

Calcula el **MNCS** de tres artículos con sus citas y la media de su campo, año y tipo.

i Ejercicio: completa los huecos _____

```r
citas <- c(12, 30, 5)
media_campo <- c(4, 40, 10)
mean(citas / ______)
```

Pista

Divide cada recuento entre la media de su campo: `media_campo`.

Solución

```
citas <- c(12, 30, 5)
media_campo <- c(4, 40, 10)
mean(citas / media_campo)
```

(3 + 0,75 + 0,5) / 3 ≈ 1,42: un 42 % por encima de la media mundial. Sin normalizar, el artículo de 30 citas parecería el mejor, pero es el de 12 citas el que destaca en su campo.

Ejercicio 8.3 · Percentiles con datos reales

¿Cuántas citas hacen falta para estar en el **10 % más citado** de la muestra?

Ejercicio: completa los huecos _____

```
unname(quantile(d$citas, ______))
```

Pista

El umbral del 10 % superior es el percentil 90: `0.9`.

Solución

```
unname(quantile(d$citas, 0.9))
```

En un estudio real, el umbral se calcula sobre **todos** los trabajos del mismo campo, año y tipo, no sobre la muestra analizada.

Actividades

Actividad: ¿Respeta el Manifiesto de Leiden?

Categorías: Lo respeta · Lo incumple

- Contratar a la persona con el índice h más alto sin leer su trabajo
- Usar PP(top 10 %) normalizado por campo junto con la valoración de un panel
- Comparar el factor de impacto medio de un grupo de historia y otro de oncología
- Permitir que los grupos evaluados revisen sus datos antes del informe
- Dar indicadores con tres decimales para distinguir a los candidatos

Solución

- Contratar a la persona con el índice h más alto sin leer su trabajo → **Lo incumple**
- Usar PP(top 10 %) normalizado por campo junto con la valoración de un panel → **Lo respeta**
- Comparar el factor de impacto medio de un grupo de historia y otro de oncología → **Lo incumple**. Ignora las diferencias entre campos.
- Permitir que los grupos evaluados revisen sus datos antes del informe → **Lo respeta**
- Dar indicadores con tres decimales para distinguir a los candidatos → **Lo incumple**. Falsa precisión.

Actividad: ¿Qué indicador es más adecuado?

Categorías: Recuento fraccionario · MNCS o percentiles · Ninguno: juicio experto

- Comparar la productividad de dos universidades con prácticas de coautoría muy distintas
- Comparar el impacto de departamentos de física y de sociología
- Valorar la originalidad de una tesis doctoral

Solución

- Comparar la productividad de dos universidades con prácticas de coautoría muy distintas → **Recuento fraccionario**
- Comparar el impacto de departamentos de física y de sociología → **MNCS o percentiles**
- Valorar la originalidad de una tesis doctoral → **Ninguno: juicio experto**

Reto evaluable

Reto del capítulo 8 · Un informe bibliométrico responsable

1. Elige una unidad (un grupo, un departamento o una revista) y descarga su producción de cinco años desde OpenAlex.
2. Calcula producción con recuento completo y fraccionario, citas normalizadas y PP(top 10 %) con los umbrales de los campos correspondientes.
3. Redacta un informe de dos páginas que siga los diez principios del Manifiesto de Leiden, incluida la explicación de las limitaciones dentro de la metodología.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Con recuento fraccionario, la suma de créditos coincide con el número de artículos. *[V / F]*
2. Un MNCS de 1 equivale a *[cero citas / la media mundial del campo / el máximo posible]*
3. La proporción de trabajos en el 10 % más citado de su campo y año se abrevia PP(top _____)
4. El factor de impacto es un indicador adecuado para evaluar artículos individuales. *[V / F]*
5. Los indicadores de percentiles son más robustos que las medias frente a valores extremos. *[V / F]*
6. Según el Manifiesto de Leiden, los indicadores deben *[sustituir al juicio experto / apoyar al juicio experto / ignorarse siempre]*

7. Contar citas en un periodo fijo tras la publicación es usar una ventana de _____
8. ¿Por qué hay que normalizar las citas por campo?

(a) Porque todos los campos citan igual (b) Porque las prácticas de citación difieren mucho entre campos y las cifras brutas no son comparables (c) Para aumentar las citas

Explicaciones

- El completo cuenta varias veces el mismo artículo.
- *Mean normalized citation score.*
- *Proportion of top 10 % publications.*
- Describe revistas, no artículos.
- Un artículo excepcional no los dispara.
- Primer principio.
- *Citation window.*
- Biomedicina frente a matemáticas.

Capítulo 9. Redes y mapas de la ciencia

Objetivos

- Distinguir las redes de citación directa, cocitación, acoplamiento bibliográfico, coautoría y coocurrencia de términos.
- Calcular medidas básicas de redes (grado, densidad, centralidad) y entender la detección de comunidades.
- Construir e interpretar mapas de la ciencia con responsabilidad.

Glosario del capítulo

Cocitación · Acoplamiento bibliográfico · Red de coautoría · Centralidad · Detección de comunidades · Mapa de la ciencia

Descargas

Plantilla de entrega del reto (.qmd) Datos de OpenAlex del capítulo (.json)

Lectura obligatoria

Boyack, K. W.; Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389-2404.

Lectura complementaria

- 2

1. Cinco formas de relacionar documentos y personas

Red	Nodos	Hay vínculo cuando…	Sirve para
Citación directa	Documentos	A cita a B	Seguir la genealogía de una idea
Cocitación (Small, 1973)	Documentos citados	C y D aparecen juntos en la bibliografía de un tercero	Ver la estructura intelectual de un campo (sus «clásicos»)
Acoplamiento bibliográfico (Kessler, 1963)	Documentos citantes	E y F comparten referencias	Detectar frentes de investigación recientes
Coautoría	Personas, instituciones, países	Firman juntos	Estudiar la colaboración
Coocurrencia de términos	Palabras o temas	Aparecen en el mismo documento	Mapear temas

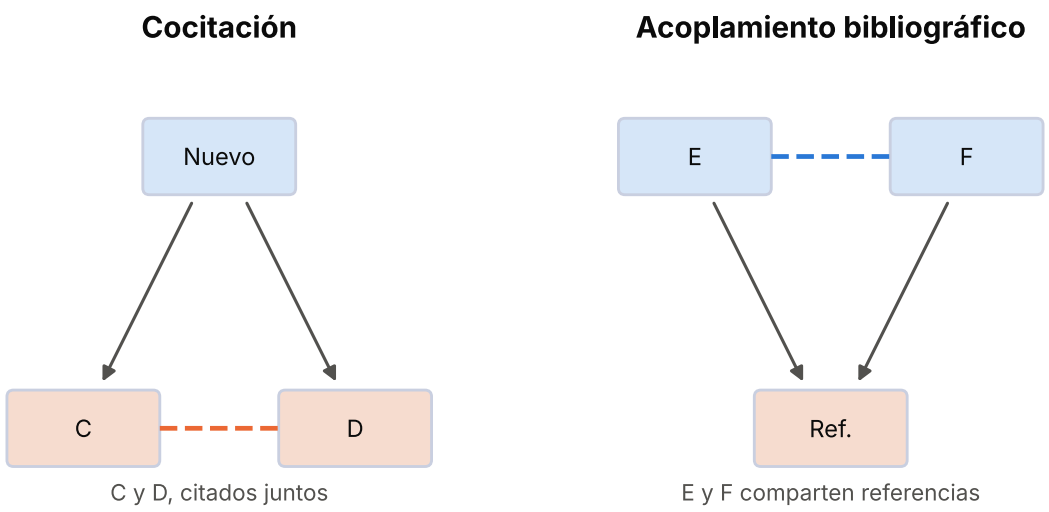

Figura 6: La cocitación une documentos citados juntos; el acoplamiento une documentos que citan lo mismo.

Boyack y Klavans (2010) compararon los tres enfoques basados en citas con más de dos millones de artículos biomédicos. Encontraron que el **acoplamiento bibliográfico** delimitaba los frentes de investigación con algo más de precisión que la cocitación, y que la citación directa era la menos precisa en ventanas cortas; los enfoques híbridos que combinan citas y texto mejoraban los resultados.

2. Medidas básicas

- **Grado**: número de vínculos de un nodo. En coautoría, con cuántas personas ha firmado alguien.
- **Peso**: intensidad del vínculo (número de referencias compartidas, de artículos conjuntos).
- **Densidad**: proporción de vínculos existentes sobre los posibles.
- **Centralidad de intermediación**: cuántos caminos más cortos pasan por un nodo; identifica «puentes» entre grupos.
- **Detección de comunidades**: algoritmos (Louvain, Leiden) que agrupan nodos más conectados entre sí que con el resto; en mapas de la ciencia, cada comunidad suele interpretarse como un tema.
- **Normalización de similitud**: coseno de Salton o fuerza de asociación, para que los documentos con bibliografías largas no dominen.

3. Mapas de la ciencia

Herramientas como **VOSviewer** (Eck y Waltman, 2010), CiteSpace, bibliometrix (R) o Gephi convierten las redes en mapas. Al leerlos, recuerda que:

- La **posición** refleja similitud, pero la distancia exacta depende del algoritmo y de sus parámetros.
- Los **umbrales** (mínimo de citas o de coocurrencias) cambian qué aparece; documenta y prueba varios.
- Los nombres de los grupos son **interpretaciones** del analista; valídalos con expertos o con lectura de muestras.
- Un mapa es una vista de una fuente y un periodo: no es «el» campo.

Ejemplo resuelto: acoplamiento con datos reales

```
d <- jsonlite::fromJSON("datos/metaciencia_openalex.json", simplifyVector =
FALSE)$trabajos
refs <- lapply(d, function(w) unlist(w$referencias))
n <- length(refs)
acopl <- matrix(0, n, n)
for (i in 1:(n - 1)) for (j in (i + 1):n)
  acopl[i, j] <- acopl[j, i] <- length(intersect(refs[[i]], refs[[j]]))
```

```r
par_max <- which(acopl == max(acopl), arr.ind = TRUE)[1, ]
cat("Mayor acoplamiento:", max(acopl), "referencias compartidas\n")
cat(" -", d[[par_max[1]]]$titulo, "\n -", d[[par_max[2]]]$titulo, "\n")
c(densidad = round(mean(acopl[upper.tri(acopl)] > 0), 2))
```

Ejercicios

Ejercicio 9.1 · Coautoría

En la red de coautoría de la muestra, ¿cuántos **coautores distintos** tiene el autor más conectado (su grado)?

i Ejercicio: completa los huecos `_____`

```r
coautores <- list()
for (w in d) {
  ids <- unique(sapply(w$autores, `[[`, "id"))
  for (a in ids) coautores[[a]] <- union(coautores[[a]], setdiff(ids, a))
}
grado <- sapply(coautores, length)
_____(grado)
```

i Pista

El grado máximo se obtiene con `max()`.

Solución

```r
coautores <- list()
for (w in d) {
  ids <- unique(sapply(w$autores, `[[`, "id"))
  for (a in ids) coautores[[a]] <- union(coautores[[a]], setdiff(ids, a))
}
grado <- sapply(coautores, length)
max(grado)
```

Los artículos con muchísimos autores (los grandes proyectos de replicación) disparan el grado: por eso a menudo se analizan aparte o se fracciona su peso.

Ejercicio 9.2 · Similitud del coseno

Dos artículos comparten 6 referencias; uno tiene 20 y otro 45. Calcula la **similitud del coseno de Salton**: compartidas / $\sqrt{(n_1 \cdot n_2)}$.

i Ejercicio: completa los huecos `_____`

```r
compartidas <- 6; n1 <- 20; n2 <- 45
round(compartidas / _____(n1 * n2), 3)
```

Pista

La raíz cuadrada en R es `sqrt()`.

Solución

```
compartidas <- 6; n1 <- 20; n2 <- 45
round(compartidas / sqrt(n1 * n2), 3)
```

0,2. La normalización evita que dos artículos con bibliografías enormes parezcan muy similares solo por citar mucho.

Actividades

Actividad: ¿Qué tipo de red necesitas?

Categorías: Cocitación · Acoplamiento bibliográfico · Coautoría · Coocurrencia de términos

- Identificar los trabajos fundacionales de la metaciencia
- Agrupar artículos publicados este año según sus frentes de investigación
- Estudiar si la colaboración internacional ha aumentado
- Ver qué temas aparecen juntos en los resúmenes

Solución

- Identificar los trabajos fundacionales de la metaciencia → **Cocitación**
- Agrupar artículos publicados este año según sus frentes de investigación → **Acoplamiento bibliográfico**
- Estudiar si la colaboración internacional ha aumentado → **Coautoría**
- Ver qué temas aparecen juntos en los resúmenes → **Coocurrencia de términos**

Actividad: ¿Qué medida de red?

Categorías: Grado · Intermediación · Densidad · Comunidades

- La investigadora que conecta dos grupos que no colaboran entre sí
- Con cuántos coautores distintos ha firmado alguien
- Qué proporción de posibles colaboraciones se han producido
- Agrupar los artículos en temas

Solución

- La investigadora que conecta dos grupos que no colaboran entre sí → **Intermediación**
- Con cuántos coautores distintos ha firmado alguien → **Grado**
- Qué proporción de posibles colaboraciones se han producido → **Densidad**
- Agrupar los artículos en temas → **Comunidades**

Reto evaluable

! Reto del capítulo 9 · Un mapa de tu campo

1. Descarga 500-2000 artículos de un tema de tu campo desde OpenAlex.
2. Construye en VOSviewer (o con bibliometrix o igraph) un mapa de acoplamiento bibliográfico y otro de coocurrencia de términos.
3. Prueba dos umbrales distintos y explica cómo cambian los grupos.
4. Interpreta las comunidades leyendo al menos cinco artículos de cada una y discute las limitaciones del mapa en la sección de metodología.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Dos documentos están acoplados bibliográficamente cuando *[uno cita al otro / comparten referencias / son citados juntos por un tercero]*
2. Dos documentos citados juntos en la bibliografía de un tercero están _____
3. Según Boyack y Klavans (2010), la citación directa fue el enfoque más preciso para delimitar frentes de investigación. *[V / F]*
4. El número de vínculos de un nodo es su _____
5. Los nombres de los grupos de un mapa son interpretaciones del analista. *[V / F]*
6. VOSviewer es una herramienta para *[gestionar referencias / construir y visualizar mapas bibliométricos / preregistrar estudios]*
7. Cambiar el umbral mínimo de coocurrencias puede cambiar los grupos del mapa. *[V / F]*
8. ¿Para qué sirve la centralidad de intermediación?

(a) Para contar publicaciones (b) Para identificar nodos que actúan de puente entre grupos (c) Para normalizar citas

Explicaciones

- Kessler (1963).
- Small (1973).
- Fue la menos precisa en ventanas cortas.
- *Degree.*
- Valídalos con lecturas o expertos.
- 2.
- Documenta los parámetros.
- *Betweenness centrality.*

Capítulo 10. Minería de textos, modelos de temas y modelos de lenguaje

Objetivos

- Preparar textos científicos para su análisis (tokenización, palabras vacías, ponderación tf-idf).
- Conocer los modelos de temas y los enfoques basados en *embeddings* y modelos de lenguaje.
- Validar cualquier clasificación automática frente a una codificación de referencia.

Glosario del capítulo

Tokenización · tf-idf · Modelo de temas · Embedding · Validación · Precisión y exhaustividad

Descargas

Plantilla de entrega del reto (.qmd) Datos de OpenAlex del capítulo (.json)

Lectura obligatoria

Grimmer, J.; Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267-297.

1. El texto como dato

Títulos, resúmenes, textos completos, informes de revisión, convocatorias o políticas científicas son fuentes masivas para la metaciencia: permiten estudiar la evolución de los temas, la novedad, el lenguaje promocional («sin precedentes», «novedoso»), la descripción de métodos o la presencia de declaraciones de datos abiertos.

Grimmer y Stewart (2013) proponen cuatro principios que siguen vigentes:

1. **Todos los modelos cuantitativos del lenguaje son erróneos, pero algunos son útiles**: simplifican el texto (por ejemplo, ignorando el orden de las palabras).
2. **Los métodos automáticos amplían, no sustituyen, la lectura humana.**
3. **No hay un método óptimo para todo**: depende de la pregunta.
4. **Validar, validar, validar**: cualquier resultado automático debe contrastarse con juicio humano.

2. Del texto a la matriz

Paso	Qué se hace	Ejemplo
Tokenización	Dividir el texto en unidades	«open science practices» → open, science, practices
Normalización	Minúsculas, quitar puntuación, a veces lematizar	Practices → practice
Palabras vacías	Eliminar palabras muy frecuentes sin contenido	the, of, and, de, la
Matriz documento-término	Contar cada término en cada documento	Filas: artículos; columnas: términos
Ponderación tf-idf	Dar más peso a términos frecuentes en un documento pero raros en el conjunto	«preregistration» pesa más que «study»

3. Métodos

- **Diccionarios**: listas de términos definidas por el investigador (por ejemplo, términos de *hype*). Transparentes, pero hay que validarlos.
- **Aprendizaje supervisado**: se codifica a mano una muestra y un modelo aprende a clasificar el resto. Se evalúa con **precisión** (de lo que el modelo marca, cuánto es correcto) y **exhaustividad** (*recall*: de lo que hay, cuánto encuentra).
- **Modelos de temas**: LDA y STM (que permite incorporar covariables como el año o el campo) descubren grupos de palabras que coaparecen. Hay que elegir el número de temas, etiquetarlos leyendo documentos y comprobar su estabilidad.
- ***Embeddings*** y modelos de lenguaje: representan textos como vectores que capturan significado; permiten agrupar por similitud semántica (BERTopic) o clasificar con modelos de lenguaje mediante instrucciones. Son potentes, pero exigen **validación** con una muestra codificada a mano, documentar el modelo y su versión, y tener en cuenta su coste y su reproducibilidad (los modelos comerciales cambian).

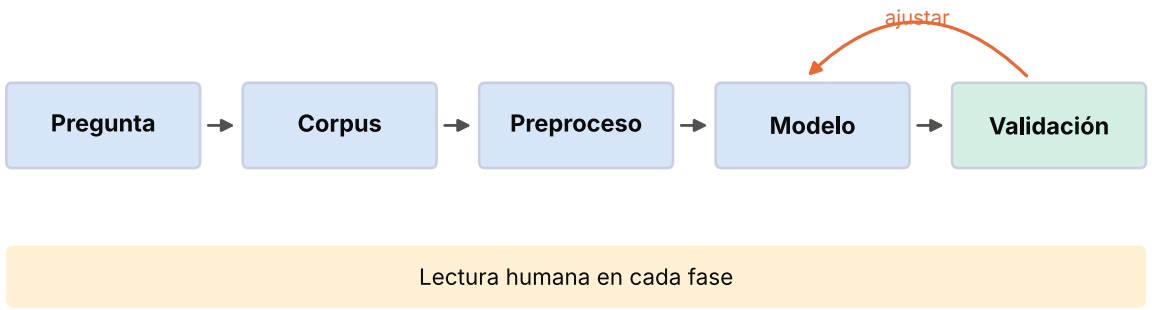

Figura 7: Flujo de un análisis de textos responsable: de la pregunta a la validación, con lectura humana en cada fase.

Ejemplo resuelto: términos más característicos de los títulos

```
d <- jsonlite::fromJSON("datos/metaciencia_openalex.json")$trabajos
vacias <-
c("the","of","and","a","in","to","for","on","with","an","is","are","from","by","at","as","how","what","
tokens <- lapply(tolower(d$titulo), function(t) {
  w <- unlist(strsplit(gsub("[^a-z ]", " ", t), " +"))
  w[nchar(w) > 2 & !w %in% vacias]
})
df <- table(unlist(lapply(tokens, unique)))     # en cuántos títulos aparece cada
término
head(sort(df, decreasing = TRUE), 12)
```

Ejercicios

Ejercicio 10.1 · tf-idf

Un término aparece 3 veces en un resumen y en 5 de 1000 documentos del corpus. Calcula su **tf-idf** con tf = 3 e idf = log(N / df).

ℹ Ejercicio: completa los huecos _____

```
tf <- 3; N <- 1000; df <- 5
round(tf * ______(N / df), 2)
```

Pista

El logaritmo natural en R es `log()`.

Solución

```
tf <- 3; N <- 1000; df <- 5
round(tf * log(N / df), 2)
```

≈ 15,9. Un término que apareciera en los 1000 documentos tendría idf = log(1) = 0: no distingue nada.

Ejercicio 10.2 · Validar un clasificador

Un modelo de lenguaje marca 40 artículos como «con datos abiertos». Al revisarlos, 30 lo son. En toda la muestra hay 50 artículos con datos abiertos. Calcula la **exhaustividad** (*recall*).

Ejercicio: completa los huecos _____

```
verdaderos_positivos <- 30; marcados <- 40; reales <- 50
precision <- verdaderos_positivos / marcados
exhaustividad <- verdaderos_positivos / ______
exhaustividad
```

Pista

La exhaustividad divide los aciertos entre todos los casos que realmente existen: `reales`.

Solución

```
verdaderos_positivos <- 30; marcados <- 40; reales <- 50
precision <- verdaderos_positivos / marcados
exhaustividad <- verdaderos_positivos / reales
exhaustividad
```

Precisión 0,75 y exhaustividad 0,6: el modelo se equivoca en una de cada cuatro marcas y se deja el 40 % de los casos. Sin esta validación, el porcentaje de datos abiertos estaría mal estimado.

Actividades

Actividad: ¿Qué método conviene?

Categorías: Diccionario · Clasificación supervisada · Modelo de temas

- Contar el uso de palabras promocionales como «novedoso» o «sin precedentes» en resúmenes
- Clasificar 100 000 artículos según declaren o no un preregistro, con 500 codificados a mano
- Explorar qué temas existen en 20 años de una revista sin categorías previas

Solución

- Contar el uso de palabras promocionales como «novedoso» o «sin precedentes» en resúmenes → **Diccionario**
- Clasificar 100 000 artículos según declaren o no un preregistro, con 500 codificados a mano → **Clasificación supervisada**
- Explorar qué temas existen en 20 años de una revista sin categorías previas → **Modelo de temas**

Actividad: Ordena un análisis con un modelo de lenguaje

- Aplicar al corpus completo y documentar todo
- Codificar a mano una muestra de referencia
- Definir la pregunta y la variable que se quiere medir
- Diseñar y fijar las instrucciones y la versión del modelo
- Medir precisión y exhaustividad frente a la muestra

Solución

1. Definir la pregunta y la variable que se quiere medir
2. Codificar a mano una muestra de referencia
3. Diseñar y fijar las instrucciones y la versión del modelo
4. Medir precisión y exhaustividad frente a la muestra
5. Aplicar al corpus completo y documentar todo

Reto evaluable

Reto del capítulo 10 · Medir algo en los resúmenes

1. Elige una característica de los textos científicos de tu campo (lenguaje promocional, declaración de datos, mención de limitaciones…).
2. Codifica tú mismo una muestra de 200 resúmenes.
3. Mídela automáticamente con dos métodos (un diccionario y un modelo supervisado o de lenguaje) y compara su precisión y exhaustividad con tu codificación.
4. Discute en la metodología las limitaciones y la reproducibilidad de cada método.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Según Grimmer y Stewart (2013), los métodos automáticos sustituyen a la lectura humana. *[V / F]*
2. Dividir un texto en unidades es la _____
3. Un término presente en todos los documentos tiene un idf de *[0 / 1 / infinito]*
4. De lo que el modelo marca, la proporción correcta es la *[exhaustividad / precisión / densidad]*
5. STM permite incorporar covariables como el año en un modelo de temas. *[V / F]*
6. Los resultados de un modelo de lenguaje comercial no necesitan validación. *[V / F]*
7. Representar textos como vectores que capturan significado es usar _____
8. ¿Qué principio resume la recomendación central de Grimmer y Stewart?

(a) Usar siempre el modelo más complejo (b) Validar, validar, validar (c) Evitar los métodos automáticos

Explicaciones

- La amplían.
- *Tokenization.*
- log(N / N) = 0.
- *Precision.*
- *Structural topic model.*
- Siempre, con una muestra codificada.
- Representaciones vectoriales.
- Cuarto principio.

Capítulo 11. Síntesis de la evidencia y detección de sesgos

Objetivos

- Diseñar revisiones sistemáticas y de alcance siguiendo PRISMA 2020.
- Calcular e interpretar un metaanálisis de efectos fijos y aleatorios y su heterogeneidad.
- Aplicar técnicas de metaciencia para detectar sesgo de publicación y errores (embudo, Egger, p-curve, GRIM, statcheck).

Glosario del capítulo

Revisión sistemática · Revisión de alcance · Metaanálisis · Heterogeneidad · Gráfico de embudo · p-curve · Prueba GRIM

Descargas

Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Page, M. J.; McKenzie, J. E.; Bossuyt, P. M.; Boutron, I.; Hoffmann, T. C. et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.

Lecturas complementarias

- Egger et al. (1997)
- Simonsohn et al. (2014)
- Brown y Heathers (2017)

1. Revisar de forma sistemática

Tipo	Pregunta	Características
Revisión narrativa	Panorama de un tema	Selección no explícita; útil para introducir, débil como evidencia
Revisión sistemática	Pregunta concreta (por ejemplo, PICO)	Protocolo previo, búsqueda exhaustiva y reproducible, criterios explícitos, evaluación del riesgo de sesgo
Revisión de alcance (*scoping*)	¿Qué se ha investigado y cómo?	Mapea la literatura; no suele evaluar calidad ni combinar efectos
Metaanálisis	¿Cuál es el efecto medio y cuánto varía?	Combina estadísticamente los resultados de los estudios
Revisión de revisiones (*umbrella*)	¿Qué dicen las revisiones existentes?	Sintetiza revisiones

PRISMA 2020 (Page et al., 2021) es la guía para **informar** de una revisión sistemática: una lista de comprobación de 27 ítems y un diagrama de flujo que muestra cuántos registros se identifican, criban, excluyen (con motivos) e incluyen. Conviene registrar el protocolo antes de empezar (PROSPERO para salud; OSF para otros campos).

2. Combinar resultados: el metaanálisis

Cada estudio aporta un tamaño del efecto (por ejemplo, *d* de Cohen) y su error típico. El **modelo de efectos fijos** supone un único efecto verdadero y pondera cada estudio por el inverso de su varianza. El **modelo de efectos aleatorios** supone que el efecto verdadero varía entre estudios (por población, diseño, contexto) y estima esa variación (τ^2).

La **heterogeneidad** se describe con Q, τ^2 e $\mathbf{I^2}$ (la proporción de la variación observada que no se debe al azar; orientativamente, 25 %, 50 % y 75 % se consideran baja, moderada y alta). Una I^2 alta pide explorar moderadores, no esconderla.

En R, el paquete **metafor** es la referencia: `rma(yi, vi, data = d)` ajusta un modelo de efectos aleatorios y `funnel()` dibuja el embudo.

3. Detectar el sesgo de publicación y los errores

Si los estudios con resultados significativos se publican más, el metaanálisis **sobrestima** el efecto. La metaciencia ha desarrollado varias herramientas:

- **Gráfico de embudo**: el efecto de cada estudio frente a su precisión. Sin sesgo, los estudios se reparten de forma simétrica; si faltan estudios pequeños con efectos nulos, aparece **asimetría**.
- **Prueba de Egger** (Egger et al., 1997): una regresión que contrasta esa asimetría. Variantes como **PET-PEESE** estiman el efecto corregido.
- **p-curve** (Simonsohn et al., 2014): analiza la distribución de los valores p significativos. Si hay un efecto real, abundan los p muy pequeños (curva sesgada a la derecha); si la curva es plana o se acumula cerca de 0,05, sugiere ausencia de efecto o prácticas cuestionables. **z-curve** extiende la idea para estimar la tasa de replicabilidad esperada.
- **GRIM** (Brown y Heathers, 2017): comprueba si una media de datos enteros (por ejemplo, de una escala Likert) es matemáticamente posible con el tamaño muestral declarado.
- **statcheck**: recalcula los valores p a partir de los estadísticos publicados (t, F, χ^2) y detecta incoherencias.

Laboratorio: un embudo con sesgo de publicación

> **i** Laboratorio interactivo
>
> Se simulan 80 estudios de un efecto verdadero. Decide qué probabilidad tienen de publicarse los estudios **no significativos** y observa el embudo y el efecto estimado con los estudios publicados.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

Ejercicios

Ejercicio 11.1 · Efectos fijos

Combina tres estudios con el **inverso de la varianza**: estimación = $\Sigma(w_i \cdot d_i) / \Sigma w_i$, con $w_i = 1 / ee_i^2$.

> **i** Ejercicio: completa los huecos _____

```r
d  <- c(0.45, 0.10, 0.30)
ee <- c(0.20, 0.08, 0.15)
w  <- 1 / ee^2
round(sum(w * d) / ______(w), 3)
```

Pista

Divide entre la suma de los pesos: `sum(w)`.

Solución

```r
d  <- c(0.45, 0.10, 0.30)
ee <- c(0.20, 0.08, 0.15)
w  <- 1 / ee^2
round(sum(w * d) / sum(w), 3)
```

≈ 0,178: el estudio más preciso (ee = 0,08) pesa más y arrastra la estimación hacia su efecto pequeño.

Ejercicio 11.2 · La prueba GRIM

Un artículo informa de una media de **3,47** en una pregunta con respuestas enteras de 1 a 7 y **n = 21**. ¿Es posible? La media de enteros debe ser un múltiplo de 1/n: comprueba si algún total entero produce 3,47 al redondear a dos decimales.

Ejercicio: completa los huecos _____

```r
n <- 21; media <- 3.47
totales <- 0:(7 * n)
any(round(totales / n, 2) == _______)
```

Pista

Compara con la media publicada: `media`.

Solución

```r
n <- 21; media <- 3.47
totales <- 0:(7 * n)
any(round(totales / n, 2) == media)
```

`FALSE`: 72/21 = 3,43 y 73/21 = 3,48; ningún total da 3,47. La media es **incoherente** con n = 21: puede ser una errata, un tamaño muestral mal informado o un problema más serio. GRIM no prueba mala conducta; indica que hay que pedir los datos.

Actividades

Actividad: ¿Qué tipo de revisión?

Categorías: Revisión sistemática con metaanálisis · Revisión de alcance · Revisión narrativa

- Estimar el efecto medio de la tutoría sobre el abandono doctoral a partir de 25 estudios
- Mapear qué métodos se han usado para estudiar la revisión por pares
- Escribir la introducción de un artículo con los trabajos más conocidos

Solución

- Estimar el efecto medio de la tutoría sobre el abandono doctoral a partir de 25 estudios → **Revisión sistemática con metaanálisis**
- Mapear qué métodos se han usado para estudiar la revisión por pares → **Revisión de alcance**
- Escribir la introducción de un artículo con los trabajos más conocidos → **Revisión narrativa**

Actividad: ¿Qué técnica detecta cada problema?

Categorías: Embudo o Egger · p-curve · GRIM · statcheck

- Faltan estudios pequeños con resultados nulos
- Los valores p significativos se acumulan justo por debajo de 0,05
- Una media de una escala de 1 a 5 imposible con el n declarado
- t(28) = 2,10 publicado con p < 0,01

Solución

- Faltan estudios pequeños con resultados nulos → **Embudo o Egger**
- Los valores p significativos se acumulan justo por debajo de 0,05 → **p-curve**
- Una media de una escala de 1 a 5 imposible con el n declarado → **GRIM**
- t(28) = 2,10 publicado con p < 0,01 → **statcheck**. Con esos grados de libertad, p ≈ 0,045.

Reto evaluable

Reto del capítulo 11 · Una revisión pequeña pero rigurosa

1. Formula una pregunta acotada de tu campo y redacta un protocolo breve (pregunta, criterios, fuentes, cadena de búsqueda).
2. Busca y criba tú mismo los resultados, documentando el diagrama de flujo PRISMA.
3. Si hay al menos cinco estudios con efectos comparables, haz un metaanálisis con metafor, dibuja el embudo y aplica la prueba de Egger. Si no, presenta una síntesis estructurada.
4. Aplica GRIM o statcheck a los estudios incluidos e informa de lo encontrado con prudencia.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. PRISMA 2020 es una guía para *[financiar revisiones / informar de revisiones sistemáticas / evaluar revistas]*
2. El modelo que supone que el efecto verdadero varía entre estudios es el de efectos _____
3. I^2 describe la proporción de variación que no se debe al azar. *[V / F]*
4. Si faltan estudios pequeños con efectos nulos, el gráfico de embudo se vuelve _____
5. Una p-curve sesgada a la derecha sugiere que hay un efecto real. *[V / F]*
6. Un resultado GRIM incoherente demuestra que hubo fraude. *[V / F]*
7. La herramienta que recalcula los valores p a partir de los estadísticos publicados es *[PRISMA / statcheck / VOSviewer]*
8. ¿Qué efecto tiene el sesgo de publicación en un metaanálisis?

(a) Ninguno (b) Tiende a sobrestimar el efecto (c) Siempre lo subestima

Explicaciones

- 27 ítems y diagrama de flujo.
- τ^2 estima la variación.
- Baja, moderada, alta: 25, 50, 75 %.
- Egger lo contrasta.
- Muchos p muy pequeños.
- Indica que hay que pedir los datos.
- t, F, χ^2.
- Faltan los resultados nulos.

Capítulo 12. Diseños de investigación sobre la investigación

Objetivos

- Elegir el diseño adecuado para preguntas causales sobre el sistema científico: experimentos, experimentos naturales y cuasiexperimentos.
- Conocer los diseños propios de la metaciencia: replicaciones a gran escala, estudios con muchos analistas y análisis multiverso.
- Valorar el papel de las encuestas, los métodos cualitativos y los métodos mixtos.

Glosario del capítulo

Inferencia causal · Experimento de campo · Diferencias en diferencias · Regresión discontinua · Muchos analistas · Análisis multiverso · Métodos mixtos

Descargas

Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Silberzahn, R.; Uhlmann, E. L.; Martin, D. P.; Anselmi, P.; Aust, F. et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337-356.

Lectura complementaria

- 3

1. Preguntas causales sobre la ciencia

Muchas preguntas de la metaciencia son causales: ¿la revisión doble ciega reduce los sesgos?, ¿las ayudas aumentan la productividad?, ¿los mandatos de datos abiertos aumentan la reutilización? Los datos bibliométricos son observacionales y están llenos de factores de confusión (los mejores grupos obtienen más fondos **y** publican más), así que hacen falta diseños que aproximen la comparación contrafactual.

Diseño	Idea	Ejemplo en metaciencia
Experimento aleatorizado	El investigador asigna al azar la intervención	Asignar al azar manuscritos a revisión simple o doble ciega
Experimento de campo	Aleatorización en un entorno real	Enviar correos idénticos con nombres distintos a investigadores para medir sesgos en la respuesta
Experimento natural	Un cambio externo asigna «casi al azar»	Una lotería entre candidatos de igual puntuación en una convocatoria
Diferencias en diferencias	Comparar el cambio antes y después en un grupo afectado frente a uno no afectado	Efecto de un mandato de acceso abierto de un financiador frente a otro que no lo aplica

Diseño	Idea	Ejemplo en metaciencia
Regresión discontinua	Comparar justo por encima y por debajo de un umbral	Proyectos que quedan justo por encima o por debajo de la nota de corte de financiación
Control sintético	Construir un «gemelo» ponderado de la unidad tratada	Un país que cambia su sistema de evaluación

Las condiciones de validez se deben argumentar: en diferencias en diferencias, las **tendencias paralelas** antes de la intervención; en la regresión discontinua, que no se pueda **manipular** la posición respecto al umbral.

2. Diseños propios de la metaciencia

- **Replicaciones a gran escala**: muchos equipos repiten estudios publicados con protocolos preregistrados (proyectos de la Open Science Collaboration, Many Labs, el proyecto de reproducibilidad en biología del cáncer). Miden la replicabilidad de un campo, no la de un estudio aislado.
- **Muchos analistas**: Silberzahn et al. (2018) dieron los mismos datos (¿reciben más tarjetas rojas los futbolistas de piel más oscura?) a 29 equipos. Los efectos estimados variaron mucho (de nulos a claramente positivos) por decisiones analíticas razonables, sin que la experiencia de los equipos explicara las diferencias. La conclusión: **la incertidumbre analítica** es una fuente de variación tan real como la muestral.
- **Análisis multiverso** (Steegen et al., 2016): en lugar de un único análisis, se ejecutan todas las combinaciones razonables de decisiones (exclusiones, codificaciones, covariables) y se informa de la distribución de resultados. Su versión gráfica es la **curva de especificaciones**.
- **Estudios de auditoría de la literatura**: codificar sistemáticamente una muestra de artículos (por ejemplo, disponibilidad de datos, preregistro, potencia) para estimar la prevalencia de prácticas.

3. Encuestas, métodos cualitativos y mixtos

No todo se mide con bases de datos. Las **encuestas** a investigadores estiman actitudes y prácticas (por ejemplo, prevalencia autodeclarada de prácticas cuestionables), con sus problemas de deseabilidad social y de respuesta. Los **métodos cualitativos** (entrevistas, grupos de discusión, etnografía de laboratorio, análisis documental de políticas y convocatorias) explican **cómo** y **por qué** funcionan las normas y los incentivos. Los **métodos mixtos** combinan ambos: por ejemplo, un análisis bibliométrico de un cambio de política seguido de análisis documental de las convocatorias para interpretar el mecanismo.



Figura 8: Tres capas de evidencia para estudiar una reforma científica: datos bibliométricos, diseños causales y comprensión cualitativa.

Laboratorio: un análisis multiverso

i Laboratorio interactivo

Una investigación estudia si los artículos con datos abiertos reciben más citas. Hay cuatro decisiones razonables con dos opciones cada una: 16 análisis posibles. Elige tu especificación favorita y compárala con el multiverso entero.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

Ejercicios

Ejercicio 12.1 · Diferencias en diferencias

Un financiador impone el acceso abierto en 2020. El porcentaje de artículos en abierto de sus proyectos pasa del 40 % al 75 %; en un financiador comparable sin mandato, del 38 % al 55 %. Calcula el efecto estimado (en puntos porcentuales).

i Ejercicio: completa los huecos _____

```
tratado_antes <- 40; tratado_despues <- 75
control_antes <- 38; control_despues <- 55
(tratado_despues - tratado_antes) - (control_despues - ______)
```

i Pista

Al cambio del grupo tratado se le resta el cambio del grupo de control: `control_despues - control_antes`.

Solución

```
tratado_antes <- 40; tratado_despues <- 75
control_antes <- 38; control_despues <- 55
(tratado_despues - tratado_antes) - (control_despues - control_antes)
```

18 puntos: de los 35 puntos de aumento, 17 habrían ocurrido igualmente (tendencia general hacia el acceso abierto). La estimación es válida si ambos grupos seguían tendencias paralelas antes de 2020.

Ejercicio 12.2 · Contar el multiverso

Un análisis tiene 3 formas razonables de tratar los valores atípicos, 2 conjuntos de covariables, 4 definiciones de la variable dependiente y 2 criterios de exclusión. ¿Cuántos análisis forman el multiverso?

Ejercicio: completa los huecos _____

```
opciones <- c(3, 2, 4, 2)
_____(opciones)
```

Pista

Las combinaciones se multiplican: `prod()`.

Solución

```
opciones <- c(3, 2, 4, 2)
prod(opciones)
```

48 análisis con solo cuatro decisiones: el «jardín de los senderos que se bifurcan» del capítulo 4 se vuelve visible.

Actividades

Actividad: ¿Qué diseño es?

Categorías: Experimento aleatorizado · Diferencias en diferencias · Regresión discontinua · Muchos analistas o multiverso

- Asignar al azar a los revisores la versión con o sin nombre de los autores
- Comparar la producción de proyectos justo por encima y por debajo de la nota de corte
- Comparar el cambio en reutilización de datos antes y después de un mandato con un grupo sin mandato
- Pedir a 20 equipos que analicen los mismos datos para responder la misma pregunta

Solución

- Asignar al azar a los revisores la versión con o sin nombre de los autores → **Experimento aleatorizado**
- Comparar la producción de proyectos justo por encima y por debajo de la nota de corte → **Regresión discontinua**
- Comparar el cambio en reutilización de datos antes y después de un mandato con un grupo sin mandato → **Diferencias en diferencias**
- Pedir a 20 equipos que analicen los mismos datos para responder la misma pregunta → **Muchos analistas o multiverso**

Actividad: ¿Método cuantitativo, cualitativo o mixto?

Categorías: Cuantitativo · Cualitativo · Mixto

- Análisis documental de 30 convocatorias para entender cómo se define la «excelencia»
- Estimar la prevalencia de preregistros en 1000 artículos codificados
- Medir el efecto de una política con datos bibliométricos y explicar su mecanismo con análisis documental

Solución

- Análisis documental de 30 convocatorias para entender cómo se define la «excelencia» → **Cualitativo**
- Estimar la prevalencia de preregistros en 1000 artículos codificados → **Cuantitativo**
- Medir el efecto de una política con datos bibliométricos y explicar su mecanismo con análisis documental → **Mixto**

Reto evaluable

Reto del capítulo 12 · Diseña tu estudio de metaciencia

1. Formula una pregunta causal sobre el sistema científico de tu entorno.
2. Elige un diseño (experimento natural, diferencias en diferencias, regresión discontinua…) que puedas llevar a cabo tú solo con datos existentes, y argumenta sus supuestos.
3. Identifica las decisiones analíticas principales y propón un análisis multiverso.
4. Redacta un preregistro breve en el formato de OSF, con las limitaciones dentro de la sección de metodología.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. En Silberzahn et al. (2018), los 29 equipos obtuvieron prácticamente el mismo efecto. *[V / F]*
2. El supuesto clave de diferencias en diferencias son las tendencias _____
3. Comparar casos justo por encima y por debajo de un umbral es una regresión *[lineal / discontinua / logística]*
4. Un análisis multiverso informa de los resultados de todas las combinaciones razonables de decisiones. *[V / F]*
5. La representación gráfica de un multiverso es la curva de _____

6. Las encuestas sobre prácticas cuestionables pueden estar afectadas por la deseabilidad social. *[V / F]*
7. Combinar datos bibliométricos con análisis documental es un diseño de métodos _____
8. ¿Qué muestran los estudios con muchos analistas?

(a) Que siempre hay un único análisis correcto (b) Que decisiones analíticas razonables pueden producir resultados muy distintos con los mismos datos (c) Que la experiencia de los equipos lo explica todo

Explicaciones

- Variaron mucho.
- Sin intervención, ambos grupos habrían cambiado igual.
- *Regression discontinuity.*
- 3.
- *Specification curve.*
- Respuestas socialmente aceptables.
- *Mixed methods.*
- Incertidumbre analítica.

Laboratorio interactivo

Este apéndice reúne laboratorios y actividades para practicar las ideas del libro con simulaciones que se ejecutan en tu navegador.

Laboratorio: la maldición del ganador y la replicación

ℹ Laboratorio interactivo

Si solo se publican los resultados significativos, los efectos publicados están **inflados**, sobre todo con muestras pequeñas. Elige el efecto real y el tamaño de muestra: el laboratorio simula 2000 estudios, «publica» los significativos y calcula qué pasaría al replicarlos.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

Actividad: ¿práctica cuestionable o correcta?

ℹ Actividad: Clasifica cada decisión

Categorías: p-hacking · HARKing · Publicación selectiva · Práctica correcta

- Probar tres formas de eliminar valores atípicos y quedarse con la que da $p < 0{,}05$
- Presentar como hipótesis previa un hallazgo descubierto al explorar los datos
- Guardar en un cajón los estudios que no salieron significativos
- Añadir participantes hasta que el resultado sea significativo
- Medir cinco resultados e informar solo del que funcionó
- Preregistrar el plan de análisis y señalar las desviaciones en el artículo
- Presentar un análisis no previsto etiquetándolo como exploratorio
- Contar después que «ya se esperaba» un efecto en un subgrupo que salió por casualidad

Solución

- Probar tres formas de eliminar valores atípicos y quedarse con la que da $p < 0{,}05$ → **p-hacking**. Decisión analítica tomada a la vista del resultado.
- Presentar como hipótesis previa un hallazgo descubierto al explorar los datos → **HARKing**. Hypothesizing After the Results are Known.
- Guardar en un cajón los estudios que no salieron significativos → **Publicación selectiva**. Es el «efecto cajón».
- Añadir participantes hasta que el resultado sea significativo → **p-hacking**. Parada opcional sin corrección.
- Medir cinco resultados e informar solo del que funcionó → **Publicación selectiva**. Notificación selectiva de resultados.
- Preregistrar el plan de análisis y señalar las desviaciones en el artículo → **Práctica correcta**. Hace visibles las decisiones.
- Presentar un análisis no previsto etiquetándolo como exploratorio → **Práctica correcta**. Es transparente sobre su naturaleza.
- Contar después que «ya se esperaba» un efecto en un subgrupo que salió por casualidad → **HARKing**. Reescribe la historia de la hipótesis.

Laboratorio: diagrama de flujo PRISMA 2020

Laboratorio interactivo

Toda revisión sistemática debe informar del recorrido de los registros, desde la búsqueda hasta los estudios incluidos, con el diagrama de **PRISMA 2020** (Page et al., 2021). Introduce tus números: el laboratorio comprueba que cuadran y dibuja el diagrama, que puedes descargar en SVG.

El diagrama simplificado sigue la versión para revisiones nuevas que solo buscan en bases de datos y registros. Si tu revisión es una actualización o usa otras vías de búsqueda (citas, webs), usa la plantilla completa de PRISMA 2020.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/ciencia-de-la-ciencia

Para ver y seguir aprendiendo

Son enlaces, no vídeos incrustados: el vídeo se abre en la web de quien lo publica, así esta página no carga rastreadores de terceros. La mayoría está en inglés; activa los subtítulos si los necesitas.

- **Is Most Published Research Wrong?** — Veritasium (2016). Unos 12 minutos sobre valores p, p-hacking y la crisis de replicación, a partir del artículo de Ioannidis (2005).

Guía para el profesorado

Este libro sirve como manual de un curso de doctorado de unas 40-48 horas sobre investigación sobre la investigación, o de dos cursos independientes: **Fundamentos** (parte I, capítulos 1-6) y **Métodos de la metaciencia** (parte II, capítulos 7-12). También puede leerse de forma autónoma. Cada capítulo equivale a una sesión de 3-4 horas con lectura previa.

La parte II usa un conjunto de datos real descargado de OpenAlex (`datos/metaciencia_openalex.json`: los 80 artículos más citados que citan a la Open Science Collaboration, 2015) para que todos los métodos se practiquen sobre los mismos datos. Los retos de la parte II están pensados para que cada estudiante pueda hacerlos solo, con datos existentes y codificación propia.

Propuesta de evaluación

Actividad	Peso orientativo	Apoyo en el libro
Retos evaluables (elegir 4 de 6)	60 %	Un reto por capítulo, con rúbrica en la plantilla de entrega
Presentación de una lectura obligatoria	20 %	Diapositivas y preguntas para debatir de cada capítulo
Participación y autoevaluaciones	20 %	Autoevaluaciones y actividades interactivas

La rúbrica común de los retos valora la pregunta y las fuentes (20 %), el rigor del análisis (30 %), la interpretación crítica (25 %), la reproducibilidad de datos y código (15 %) y la presentación (10 %). Los retos no tienen solución en el libro: cada persona trabaja sobre su propio campo.

Actividades interactivas y laboratorios

Cada capítulo incluye actividades de corrección inmediata y laboratorios que se usan sin programar: índice h, valor predictivo positivo de un resultado significativo, simulación de *p-hacking* y escenarios de decisión (análisis que no sale, planificación de prácticas abiertas, comisión de selección). Los ejercicios de R se ejecutan en el navegador con webR. En el PDF y el EPUB, las actividades aparecen en papel con su solución.

Autoevaluación en Aula Global

Las preguntas de autoevaluación están exportadas a **Moodle XML** en la carpeta `moodle/` del repositorio. Se importan desde *Banco de preguntas* → *Importar* → *Formato Moodle XML*.

Descarga el banco de preguntas: todo el libro (Moodle XML) · por capítulo: 00 · 01 · 02 · 03 · 04 · 05 · 06 · 07 · 08 · 09 · 10 · 11 · 12

Diapositivas

Cada capítulo tiene diapositivas con los objetivos, las figuras, las tablas, la lectura obligatoria, preguntas para debatir y el reto: diapositivas/ (F, pantalla completa; S, notas).

Mantenimiento

```
python3 herramientas/preparar.py        # sw.js y flujo de integración continua
python3 herramientas/extraer_r.py       # scripts .R de cada capítulo
python3 herramientas/moodle.py          # banco de preguntas
python3 herramientas/figuras.py         # figuras
python3 herramientas/pdf.py             # PDF y EPUB
quarto render
python3 herramientas/diapositivas.py    # diapositivas de clase
```

Los datos de las simulaciones son generados con semilla fija y no corresponden a personas reales.

Glosario

Términos del libro ordenados por capítulo. Cada entrada enlaza con el capítulo donde se explica.

Qué es la ciencia de la ciencia

Ciencia de la ciencia
Science of science. Estudio científico de cómo se produce, comunica y evalúa el conocimiento, con datos sobre publicaciones, citas, colaboraciones y financiación. → Capítulo 1

Metaciencia
Investigación sobre la fiabilidad de los métodos y resultados científicos y sobre las reformas que la mejoran. También, metainvestigación. → Capítulo 1

Investigación sobre la investigación
Research on research (RoR). Estudio de qué hace que la investigación sea mejor y más útil, con atención especial a la financiación y la política científica. → Capítulo 1

Cienciometría
Estudio cuantitativo de la ciencia, sobre todo a partir de publicaciones y citas. → Capítulo 1

Efecto Mateo
Acumulación de reconocimiento en quienes ya lo tienen, descrita por Merton (1968). → Capítulo 1

Ley de Price
Observación de que una pequeña parte de los autores produce una gran parte de las publicaciones de un campo. → Capítulo 1

Medir la ciencia

Cita
Referencia de un documento a otro. Mide uso o visibilidad, no calidad. → Capítulo 2

Índice h
Mayor número h tal que h trabajos tienen al menos h citas cada uno (Hirsch, 2005). → Capítulo 2

Factor de impacto
Media de citas recibidas en un año por los artículos que una revista publicó en los dos años anteriores. Es un indicador de revista. → Capítulo 2

Normalización por campo
Comparación de las citas de un trabajo con la media de los trabajos del mismo campo, año y tipo. → Capítulo 2

Distribución asimétrica
Distribución en la que pocos casos acumulan valores muy altos; la media no representa lo típico. → Capítulo 2

Altmetría
Indicadores basados en menciones en redes sociales, noticias, documentos de políticas o gestores de referencias. → Capítulo 2

Manifiesto de Leiden
Diez principios para el uso responsable de los indicadores en la evaluación (Hicks y others, 2015). → Capítulo 2

Reproducibilidad

Reproducibilidad
Obtener los mismos resultados con los mismos datos y el mismo código. → Capítulo 3

Replicabilidad
Obtener resultados coherentes con datos nuevos recogidos con el mismo método. → Capítulo 3

Potencia estadística
Probabilidad de que un estudio detecte un efecto si existe. → Capítulo 3

Valor predictivo positivo
Probabilidad de que un resultado significativo corresponda a un efecto real. → Capítulo 3

Falso positivo
Resultado significativo cuando no existe un efecto real (error de tipo I). → Capítulo 3

Maldición del ganador
Tendencia de los estudios pequeños que resultan significativos a sobrestimar el tamaño del efecto. → Capítulo 3

Prácticas cuestionables

Prácticas cuestionables
Decisiones que no son fraude pero aumentan la probabilidad de resultados falsos, como informar solo de lo significativo. → Capítulo 4

p-hacking
Probar análisis o decisiones hasta obtener un resultado significativo. → Capítulo 4

HARKing
Hypothesizing after the results are known: presentar como hipótesis previa lo descubierto tras ver los datos. → Capítulo 4

Grados de libertad del investigador
Decisiones que se pueden tomar al recoger y analizar los datos y que cambian el resultado. → Capítulo 4

Sesgo de publicación
Dependencia de la publicación respecto del resultado: se publican más los resultados positivos. → Capítulo 4

Parada opcional
Decidir cuándo dejar de recoger datos según el resultado provisional. → Capítulo 4

Reformas

Ciencia abierta
Conjunto de prácticas que hacen accesibles y transparentes las publicaciones, los datos, el código, los métodos y la evaluación. → Capítulo 5

Preregistro
Plan del estudio registrado con fecha antes de analizar los datos. → Capítulo 5

Informe registrado
Formato de publicación en el que la revista acepta el estudio antes de conocer los resultados. → Capítulo 5

Datos FAIR
Datos localizables, accesibles, interoperables y reutilizables. → Capítulo 5

Insignias de prácticas abiertas
Distintivos que algunas revistas conceden a los artículos que comparten datos, materiales o preregistros. → Capítulo 5

Preprint
Versión de un trabajo difundida antes de la revisión por pares. → Capítulo 5

Evaluación

Evaluación responsable
Evaluación basada en el juicio experto, apoyada por indicadores bien usados y coherente con la misión de lo evaluado. → Capítulo 6

CoARA
Coalición para el avance de la evaluación de la investigación, creada en 2022 en torno a un acuerdo europeo con diez compromisos. → Capítulo 6

DORA
Declaración de San Francisco sobre la evaluación de la investigación (2012). → Capítulo 6

Principios de Hong Kong
Cinco principios para evaluar a las personas investigadoras fomentando la integridad (Moher et al., 2020). → Capítulo 6

CV narrativo
Currículo que presenta contribuciones seleccionadas y explicadas en lugar de listas exhaustivas. → Capítulo 6

Revisión por pares
Evaluación de un trabajo o una propuesta por especialistas del área. → Capítulo 6

CRediT
Taxonomía de 14 roles para declarar la contribución de cada persona a un trabajo. → Capítulo 6

Fuentes de datos

Base de datos bibliográfica
Fuente que indexa publicaciones con sus metadatos y, a menudo, sus citas (Web of Science, Scopus, Dimensions, OpenAlex). → Capítulo 7

Cobertura
Conjunto de documentos, campos, idiomas y años que recoge una fuente. → Capítulo 7

Identificador persistente
Código estable que identifica sin ambigüedad un objeto, persona u organización (DOI, ORCID, ROR). → Capítulo 7

Desambiguación de autores
Asignación correcta de las publicaciones a cada persona pese a homónimos y variantes del nombre. → Capítulo 7

Metadatos abiertos
Metadatos de la investigación que cualquiera puede reutilizar sin restricciones. → Capítulo 7

Bibliometría e indicadores

Recuento completo
Cada unidad que participa en un trabajo suma una publicación completa. → Capítulo 8

Recuento fraccionario
El trabajo se reparte entre las unidades participantes, de modo que la suma es uno. → Capítulo 8

MNCS
Media de las citaciones normalizadas por campo, año y tipo de documento de un conjunto de trabajos. → Capítulo 8

Indicador de percentil
Proporción de trabajos de una unidad que están entre los más citados de su campo y año, como PP(top 10 %). → Capítulo 8

Ventana de citación
Periodo fijo tras la publicación en el que se cuentan las citas. → Capítulo 8

Redes y mapas

Cocitación
Relación entre dos documentos citados juntos por un tercero. → Capítulo 9

Acoplamiento bibliográfico
Relación entre dos documentos que comparten referencias. → Capítulo 9

Red de coautoría
Red en la que personas, instituciones o países se unen al firmar juntos. → Capítulo 9

Centralidad
Medida de la importancia de un nodo en una red (grado, intermediación…). → Capítulo 9

Detección de comunidades
Algoritmos que agrupan los nodos más conectados entre sí (Louvain, Leiden). → Capítulo 9

Mapa de la ciencia
Representación visual de la estructura de un campo a partir de redes bibliométricas. → Capítulo 9

Textos

Tokenización
División de un texto en unidades de análisis. → Capítulo 10

tf-idf
Ponderación que valora los términos frecuentes en un documento pero raros en el corpus. → Capítulo 10

Modelo de temas
Modelo que descubre grupos de palabras que coaparecen (LDA, STM). → Capítulo 10

Embedding
Representación de un texto como vector numérico que captura su significado. → Capítulo 10

Validación
Contraste de una medida automática con una codificación humana de referencia. → Capítulo 10

Precisión y exhaustividad
Proporción de aciertos entre lo marcado (precisión) y entre lo existente (exhaustividad). → Capítulo 10

Síntesis de la evidencia

Revisión sistemática
Revisión con protocolo, búsqueda exhaustiva y criterios explícitos. → Capítulo 11

Revisión de alcance
Revisión que mapea qué se ha investigado y cómo (scoping review). → Capítulo 11

Metaanálisis
Combinación estadística de los resultados de varios estudios. → Capítulo 11

Heterogeneidad
Variación de los efectos entre estudios más allá del azar (τ^2, I^2). → Capítulo 11

Gráfico de embudo
Gráfico del efecto de cada estudio frente a su precisión, para detectar asimetrías. → Capítulo 11

p-curve
Análisis de la distribución de los valores p significativos para valorar si hay un efecto real. → Capítulo 11

Prueba GRIM
Comprobación de si una media de datos enteros es posible con el tamaño muestral declarado. → Capítulo 11

Diseños de investigación

Inferencia causal
Estimación del efecto de una intervención frente a lo que habría ocurrido sin ella. → Capítulo 12

Experimento de campo
Experimento aleatorizado en un entorno real. → Capítulo 12

Diferencias en diferencias
Comparación del cambio antes y después entre un grupo afectado y otro no afectado. → Capítulo 12

Regresión discontinua
Comparación de casos justo a ambos lados de un umbral de asignación. → Capítulo 12

Muchos analistas
Diseño en el que varios equipos analizan los mismos datos para responder la misma pregunta. → Capítulo 12

Análisis multiverso
Ejecución de todas las combinaciones razonables de decisiones analíticas. → Capítulo 12

Métodos mixtos
Combinación de métodos cuantitativos y cualitativos en un mismo estudio. → Capítulo 12

Referencias

Bibliografía

Boyack, K. W., y Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?. *Journal of the American Society for Information Science and Technology*, *61*(12), 2389-2404. https://doi.org/10.1002/asi.21419

Brown, N. J. L., y Heathers, J. A. J. (2017). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science*, *8*(4), 363-369. https://doi.org/10.1177/1948550616673876

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., y Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376. https://doi.org/10.1038/nrn3475

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at \emph{Cortex}. *Cortex*, *49*(3), 609-610. https://doi.org/10.1016/j.cortex.2012.12.016

Eck, N. J. van, y Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

Egger, M., Davey Smith, G., Schneider, M., y Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Fanelli, D. (2010). «Positive» results increase down the hierarchy of the sciences. *PLoS ONE*, *5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., y others. (2018). Science of science. *Science*, *359*(6379), eaao185. https://doi.org/10.1126/science.aao0185

Grimmer, J., y Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267-297. https://doi.org/10.1093/pan/mps028

Hicks, D., y others. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*, 429-431. https://doi.org/10.1038/520429a

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS*, *102*(46), 16569-16572. https://doi.org/10.1073/pnas.0507655102

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., y Goodman, S. N. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLOS Biology*, *13*(10), e1002264. https://doi.org/10.1371/journal.pbio.1002264

John, L. K., Loewenstein, G., y Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524-532. https://doi.org/10.1177/0956797611430953

Kaplan, R. M., y Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS ONE*, *10*(8), e132382. https://doi.org/10.1371/journal.pone.0132382

Kidwell, M. C., y others. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56-63. https://doi.org/10.1126/science.159.3810.56

Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., y others. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, *18*(7), e3000737. https://doi.org/10.1371/journal.pbio.3000737

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. Percie du, y others. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 21. https://doi.org/10.1038/s41562--016-0021

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., y Mellor, D. T. (2018). The preregistration revolution. *PNAS*, *115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., y others. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Price, D. J. d. S. (1965). Networks of scientific papers. *Science*, *149*(3683), 510-515. https://doi.org/10.1126/science.149.3683.510

Scheel, A. M., Schijen, M. R. M. J., y Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science*, *4*(2). https://doi.org/10.1177/25152459211007467

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., y others. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337-356. https://doi.org/10.1177/2515245917747646

Simmons, J. P., Nelson, L. D., y Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., y Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534-547. https://doi.org/10.1037/a0033242

Sinatra, R., Wang, D., Deville, P., Song, C., y Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312), aaf5239. https://doi.org/10.1126/science.aaf5239

Smaldino, P. E., y McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384. https://doi.org/10.1098/rsos.160384

Steegen, S., Tuerlinckx, F., Gelman, A., y Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702-712. https://doi.org/10.1177/1745691616658637

Visser, M., Eck, N. J. van, y Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20-41. https://doi.org/10.1162/qss_a_00112

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365-391. https://doi.org/10.1016/j.joi.2016.02.007

Wang, D., Song, C., y Barabási, A.-L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127-132. https://doi.org/10.1126/science.1237825

Wuchty, S., Jones, B. F., y Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036-1039. https://doi.org/10.1126/science.1136099