

Instituto de Cálculo

**Seminario en la Interfase
Diez Pautas Simples para una Práctica
Efectiva de la Estadística**

Laura Ación
en colaboración con L. Paloma Rojas Saunero



laura.acion@ic.fcen.uba.ar



[@_lacion_](https://twitter.com/_lacion_)

EXACTAS UBA

Consecuencias de un Mal Uso de la Estadística

ORIGINAL CONTRIBUTION

JAMA, July 13, 2005—Vol 294, No. 2

Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

PLoS Medicine | www.plosmedicine.org

August 2005 | Volume 2 | Issue 8 | e124

Why Most Published Research Findings Are False

John P. A. Ioannidis

Clinical Chemistry 57:5
688–690 (2011)

PNAS

7900–7905 | PNAS | July 12, 2016 | vol. 113 | no. 28

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

What Information Should Be Required to Support Clinical “Omics” Publications?

Results of Many High-Throughput “Omics” Studies Are Difficult to Reproduce, in Part Because the Data and Methods Supplied Are Inadequate



ELSEVIER

Journal of
Clinical
Epidemiology

Journal of Clinical Epidemiology ■ (2016) ■

Evidence-based medicine has been hijacked

Consecuencias de un Mal Uso de la Estadística

Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab

Tim van der Zee^{*1}, Jordan Anaya^{*2}, Nicholas J L Brown^{*3}

January 25, 2017

NOT PEER-REVIEWED
"PeerJ Preprints"

- ¡4 papers con más de 150 errores!
 - Detectados sin tener acceso a los datos originales.
- El investigador principal tiene unas 500 publicaciones y es un investigador muy respetado y con mucha llegada en los medios en el área de nutrición.
- La saga de la pizza continua: ya son 45 los papers con problemas (Anaya et al, 2017).

Diez Pautas Simples



EDITORIAL

Ten Simple Rules for Effective Statistical Practice

Robert E. Kass¹, Brian S. Caffo², Marie Davidian³, Xiao-Li Meng⁴, Bin Yu⁵, Nancy Reid^{6*}

Dirigido a investigadores con algún conocimiento de estadística:

- Con la posibilidad de consultar con un estadístico o
- Con la sana actitud de hacer el análisis estadístico por sí mismos

Pauta 1

“Los métodos estadísticos deben permitir que los datos respondan preguntas científicas”.

- Usuario de estadística sin experiencia:
 - Da por sentada la conexión entre los datos y la pregunta.
 - Busca la herramienta estadística que se aplica a la estructura de datos, aplica la estadística como una receta.
- Estadístico experimentado:
 - Estudia la pregunta y su contexto, discute con los investigadores cómo los datos podrían contestar la pregunta y qué diseños podrían ser más útiles.
 - Busca las fuentes de variabilidad potenciales, las realidades ocultas que podrían romper la relación entre los datos y la inferencia.
 - Luego decide qué herramienta analítica es más apropiada.

Pauta 3

“Planifique su estudio con tiempo, con mucho tiempo”.

- Cuando se va a hacer un gran esfuerzo colectando datos, los problemas estadísticos no se restringen únicamente a la pregunta “¿qué tamaño de muestra necesito?” (<https://www.youtube.com/watch?v=Hz1fyhVOjr4>).
- Ante esa pregunta, un estadístico experimentado va a necesitar varias iteraciones (= tiempo, mucho tiempo) con el investigador para poder contestarla.
 - ¿Cuál es la variable respuesta ideal?, ¿cómo se interpreta?, ¿cómo se miden las variables?, ¿qué factores alteran la medición?, ¿hay fuentes de error que se puedan controlar?, ¿hay sesgos en la medición?, etc.
 - Una colaboración estadístico-investigador funciona cuando genera puntos de vista complementarios que mejoran la calidad del proyecto.

Pauta 4

“Pre/ocúpese por la calidad de los datos”.

- Malos datos = malos resultados (garbage in, garbage out - aunque tenga muchos datos/big data).
- La estadística tiene herramientas que a veces pueden ayudar, pero nada funciona tan bien como datos de buena calidad.
- Los datos que vienen pre-procesados por cajas negras (por ej. genética, imágenes) requieren especial cuidado.
- El proceso de limpieza de datos (idealmente durante la ejecución del estudio) es particularmente importante porque ayuda a detectar problemas en los datos. Es imprescindible:
 - Saber cómo se colectaron los datos y qué significan (celdas vacías, códigos 999, etc).
 - Describir bien los datos mediante estadísticos descriptivos y gráficos, conocerlos, explorarlos (detección de outliers y otros problemas).

Pauta 5

“El análisis estadístico es mucho más que una serie de cálculos matemáticos”.

- Los paquetes estadísticos proveen herramientas para asistir en el análisis, pero no definen qué análisis es adecuado.
- Los métodos analíticos deben tener una correspondencia cercana con las preguntas científicas.
 - Hay que planearlos **antes** de obtener los datos e incluirlos en el protocolo de investigación (¡también si el estudio es observacional!).
- Decir qué paquete o función estadística se usó en los materiales y métodos no reemplaza explicar el razonamiento detrás de usar una herramienta estadística para responder una pregunta.

Pauta 6

“Mantenga la simplicidad”.

- Principio de parsimonia: Navaja de Occam, KISS, menos es más, la simplicidad es la máxima sofisticación.
- Sólo agregar complejidad al análisis si es necesario y en la menor medida posible.
- Sin embargo, los datos científicos pueden tener una estructura compleja que no siempre es bien modelada por modelos simples.
 - Falta de independencia, interacciones entre predictores, mecanismos no lineales, datos faltantes, confundidores, sesgos, etc. pueden requerir modelos complejos.
- Un diseño bueno y bien implementado puede contribuir a un análisis más simple y producir resultados contundentes.

Pauta 7

“Informe una evaluación de la variabilidad”.

- Cuando se reportan resultados es esencial suministrar alguna noción de la incerteza estadística.
 - Con el valor p **no** alcanza.
 - **Siempre que se pueda**, hay que usar intervalos de confianza.
- Si uno repite el experimento, incluso en la misma muestra de individuos, los datos van a cambiar.
 - Más aún si se trata de una muestra nueva o si se hacen modificaciones al protocolo (aunque sean mínimas).
- Si la muestra es muy grande en algún sentido (grandes datos) esto también se aplica.

REITERAMOS

crónica

16:21

26:5

ULTIMO MOMENTO

**Statisticians issue
warning on *P* values**

Valor P y Controversias Asociadas

Declaración de la Asociación de Estadística Americana sobre la Significación Estadística y los Valores P

- “1. Los valores p pueden indicar qué tan incompatibles son los datos con una hipótesis o un modelo estadístico.”
- Cuanto menor es el valor p , mayor será la incompatibilidad estadística de los datos con la hipótesis nula, **siempre y cuando los supuestos subyacentes sean adecuados.**
 - La incompatibilidad puede ser interpretada como un elemento que genera duda sobre la veracidad, o proveer evidencia en contra, de la hipótesis nula **o los supuestos.**

Pauta 8

“Chequee los supuestos”.

- **Todas** las técnicas estadísticas tienen supuestos, incluso las no paramétricas.
- Es necesario entender los supuestos de cada herramienta estadística que se usa y chequearlos antes de sacar conclusiones.
- Como mínimo, siempre hay que chequear cuán bien se ajusta el modelo impuesto a los datos (bondad de ajuste).
 - Usando gráficos de residuos
 - Usando pruebas estadísticas de bondad de ajuste

Valor P y Controversias Asociadas

“3. Las conclusiones científicas y las decisiones políticas o económicas **no deben** basarse únicamente en el valor p.”

- Reducir el análisis de datos a reglas mecánicas (como $p < 0,05$) puede llevar a conclusiones y decisiones incorrectas.
- Una conclusión no se convierte en “verdadera” de un lado de 0,05 y en “falsa” del otro lado.
- Las conclusiones requieren un contexto que incluye:
 - Diseño del estudio
 - Calidad de las mediciones
 - Evidencia externa sobre el fenómeno estudiado
 - Validez de los supuestos del análisis estadístico
- El uso de la significación estadística ($p < 0,05$) para afirmar un descubrimiento científico lleva a una distorsión considerable del proceso científico.

Valor P y Controversias Asociadas

“6. Por sí sólo, un valor p no proporciona una buena medida de la evidencia con respecto a una hipótesis.”

- Un valor p fuera de contexto o sin otra evidencia brinda información limitada.
 - Un valor p cercano a 0,05 tomado aisladamente ofrece evidencia débil en contra de la hipótesis nula.
 - Un valor p relativamente grande no implica evidencia a favor de la hipótesis nula – muchas otras hipótesis pueden ser igual o más consistentes con los datos observados.
- El análisis no debe terminar con el cálculo de un valor p cuando otros enfoques son factibles (por ej., ICs).

Pauta 9

“Siempre que sea posible, ¡replique!”.

- Todo buen analista examina mucho los datos en busca de patrones de muchos tipos y pone a prueba resultados tanto esperados como no esperados.
- Este proceso suele involucrar docenas de pruebas estadísticas.
- Eventualmente alguna característica de los datos resulta interesante y esto es lo que se suele reportar como resultado.
- Cuando se reporta un valor p que proviene de haber examinado los datos extensivamente y no se reportan todos los análisis realizados es análogo a:

Hacer trampa



Valor P y Controversias Asociadas

“4. Realizar inferencias adecuadas requiere informes completos y transparentes.”

- Los valores p y los análisis relacionados no deberían ser reportados en forma selectiva.
- Conducir análisis múltiples de los datos y reportar sólo aquéllos que son estadísticamente significativos no permite interpretar los p reportados correctamente.
 - Otros nombres: *cherry-picking*, *dragado de datos*, *caza/búsqueda de significación*, *p hacking*, etc.
- Estas prácticas llevan a un exceso de resultados espurios en la literatura y deben ser enfáticamente evitados.

Valor P y Controversias Asociadas

“4. Realizar inferencias adecuadas requiere informes completos y transparentes.”

- Cada vez que un investigador elige qué resultados presentar basándose en resultados estadísticos y no comunica y justifica su elección, es muy difícil que el lector haga una interpretación correcta de los resultados.
- Los investigadores deben informar: **número de hipótesis exploradas** durante el estudio, **todas** las decisiones acerca de la recolección de datos, **todos** los análisis realizados y **todos** los valores p calculados.
- Sin esta información, no se debería enunciar conclusiones basadas en valores p.

Pauta 9: “Siempre que sea posible, ¡replique!”

Algunas posibles soluciones (aún no hay consenso experto):

- Ajustar por comparaciones múltiples cuando se hicieron muchos análisis.
- Tener un plan analítico basado en hipótesis a priori registrado **antes** de coleccionar los datos (información útil sobre pre-registro de estudios de todo tipo: <http://www.timvanderzee.com/what-is-pre-registration/>).
- Replicar el resultado:
 - Muestra nueva
 - Experimento nuevo
 - Investigador independiente
- Los resultados científicos que resisten el paso del tiempo son aquéllos que son confirmados a través de una variedad de situaciones diferentes pero relacionadas.

Pauta 10

“Haga análisis reproducibles”.

La replicación casi nunca es realizable. Estándar mínimo: que los resultados sean reproducibles.

- Reproducibilidad: dados los mismos datos y la descripción completa del análisis se debe poder reproducir tablas, figuras e inferencias estadísticas.
- Herramientas para lograrlo:
 - Análisis protocolizado **antes** de coleccionar los datos.
 - Uso de programación para el análisis (por ej, R).
 - Publicar los datos y el código estadístico usados para producir los resultados reportados.
 - En R, Sweave y knitr combinan reportes con código usado para procesar los datos (existen otras herramientas fuera de R).
 - Ver Goldman et al (2014). Ten simple rules for the care and feeding of scientific data. PLoS Comput Biol 10(4):e1003542.

Valor P y sus Controversias un Año Después

- Es cierto. No hay una forma única y perfecta para transformar los datos en conocimiento.
- ¡Es sorprendente que alguien crea que tal cosa existe!
- La Ciencia es compleja.
- Realizar inferencias es un trabajo difícil.
- Ha sido extraordinariamente costoso para la Ciencia que generaciones enteras de investigadores consideren que un valor p , o cualquier otro índice, podría proveer una respuesta simple, clara y objetiva a la pregunta “¿qué nos dicen estos datos?”

Valor P y sus Controversias un Año Después



Mesa redonda "ASA statement about p-values: What now?"

XIV Congreso Latinoamericano de Probabilidad y Estadística Matemática. Diciembre 2016.

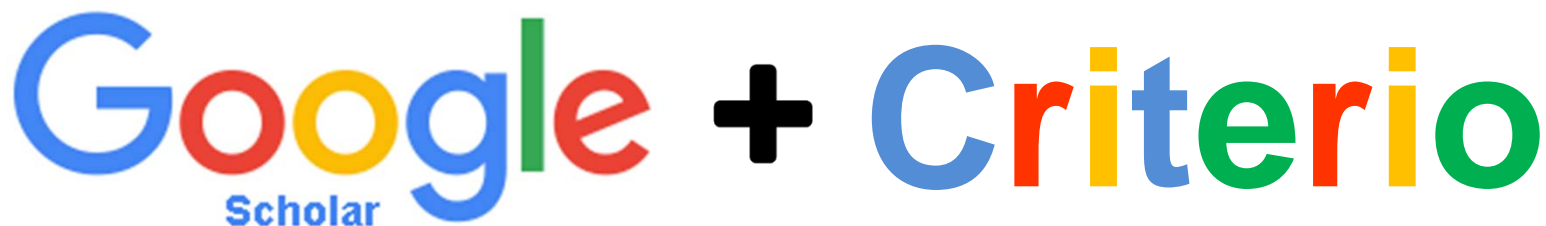


Pauta 0

Regla de Oro de la Buena Práctica Estadística

**La Estadística es una Ciencia,
no es sólo una Técnica,
ni debe ser usada como una receta.**

Con ustedes, sus mejores amigos para mantenerse al día de qué constituye la buena práctica estadística:



Referencias

- Anaya, J., van der Zee, T., Brown N. (2017) Statistical infarction: A postmortem of the Cornell Food and Brand Lab pizza publications. PeerJ Preprints 5:e3025v1
<https://doi.org/10.7287/peerj.preprints.3025v1>
- Baggerly, K., Coombes, K. What information should be required to support clinical “omics” publications? *Clinical chemistry* 57.5 (2011): 688-690.
- Baker, M. Statisticians issue warning on P values. Statement aims to halt missteps in the quest for certainty. *Nature* 531 (2016):151.
- Eklund, A., Nichols, T., Knutsson, H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences* (2016): 201602413.
- Ioannidis, J. Contradicted and initially stronger effects in highly cited clinical research. *Jama* 294.2 (2005): 218-228.
- Ioannidis, J. Why most published research findings are false. *PLoS Med* 2.8 (2005): e124.
- Ioannidis, J. Evidence-based medicine has been hijacked: a report to David Sackett. *Journal of clinical epidemiology* 73 (2016): 82-86.
- **Kass, R. et al. Ten Simple Rules for Effective Statistical Practice. *PLOS Comput Biol* 12.6 (2016): e1004961.**
- Matthews, R., Wasserstein, R. and Spiegelhalter, D. (2017), The ASA's *p*-value statement, one year on. *Significance*, 14: 38–41. doi:10.1111/j.1740-9713.2017.01021.x
- van der Zee T., Anaya J., Brown N. (2017) Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab. PeerJ Preprints 5:e2748v1
<https://doi.org/10.7287/peerj.preprints.2748v1>
- **Wasserstein, R., N. Lazar. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 70.2 (2016): 129-133. Original en inglés y también en español.**