

## Rompiendo paradigmas en el concepto de Prueba de Hipótesis utilizando tecnología

## Breaking paradigms in Hypothesis Test concept through technology

**Lino González, Montserrat**

Universidad Autónoma de Querétaro

[montserrat.lino@uaq.mx](mailto:montserrat.lino@uaq.mx)



0000-0002-6412-7890

**Chaparro Sánchez, Ricardo**

Universidad Autónoma de Querétaro

[rchapa@uaq.mx](mailto:rchapa@uaq.mx)



0000-0002-6842-2360

**Ruiz Velasco-Acosta, Silvia**

Universidad Nacional Autónoma de México

[silvia@sigma.iimas.unam.mx](mailto:silvia@sigma.iimas.unam.mx)



0000-0002-9220-2368

Sección: **Ensayo científico**

# Resumen

En libros y cursos de estadística se suele enseñar el método inferencial de la prueba de hipótesis como la unión de los métodos de Fisher y Neyman-Pearson. La conjunción de estos dos métodos, creados con distinto propósito, genera dificultades epistemológicas en su comprensión y aplicación, haciéndola susceptible de conclusiones e interpretaciones incorrectas al apoyarse únicamente en la significación estadística. Es un hecho conocido que para un tamaño de muestra grande todas las hipótesis se rechazan. Se podría elaborar una herramienta que encuentre, en una gran base de datos, el tamaño de muestra a partir del cual se empieza a rechazar la hipótesis nula, con el propósito de hacer conciencia en los estudiantes de la importancia del tamaño del efecto como elemento necesario, aunado a los intervalos de confianza, potencia, al estadístico de prueba y el valor  $p$  que lo acompaña, para tomar una decisión. Se lograría así una cultura de utilizar y reportar varias herramientas para justificar las conclusiones e importancia de los hallazgos encontrados en un estudio evitando cometer errores de tipo I al aumentar el tamaño de la potencia debido al aumento del tamaño de muestra en grandes bases de datos.

**Palabras clave:** Rechazo hipótesis nula; error tipo I; tamaño del efecto, simulación; grandes bases de datos

# Abstract

In statistics books and courses, the inferential method of hypothesis testing is usually taught, such as the union of the Fisher and Neyman-Pearson methods. The conjunction of these two methods created for different purposes, generates epistemological difficulties in its understanding and application, making it susceptible to incorrect conclusions and interpretations by only relying on statistical significance. It is a fact known that for a large sample size all hypotheses are rejected. A tool could be developed that finds, in a large

database, the sample size from which the null hypothesis begins to be rejected, to make students aware of the importance of effect size as a necessary element, together with the confidence intervals, power, test statistic and accompanying p-value, to take a decision. With this, it would be reached a culture of using and reporting several tools to justify the conclusions and importance of the findings found in a study, avoiding type I errors when increasing the power size due to the rise in the sample size in large databases.

**Keywords:** Null hypothesis rejected; type I error; effect size; simulation; large databases.

## 1. Introducción

La estadística es una asignatura obligatoria en la mayoría de las licenciaturas e ingenierías. La inferencia estadística, que utiliza las pruebas de hipótesis, es uno de los temas importantes a analizar por las dificultades epistemológicas en su comprensión y aplicación.

En cursos y libros de estadística se suele enseñar el método inferencial de la prueba de hipótesis como la unión de los métodos de Fisher y Neyman-Pearson obteniendo conclusiones e interpretaciones erróneas al basarse únicamente en la significación estadística.

Resultado de esta enseñanza se ve reflejada en publicaciones de investigación donde, a través de pruebas de significancia estadísticas, al aplicar algún estadístico de prueba como medio para rechazar o no la hipótesis nula determinan la validez de sus resultados (Leenen, 2012). Sin embargo, no se suelen enseñar otros conceptos como el tamaño del efecto que evitan cometer un error de tipo I, rechazar una hipótesis nula verdadera, pues ayudan a interpretar el significado de los resultados y ver realmente que tan grande o fuerte es la diferencia con la que estamos rechazando una hipótesis nula. Esto es importante, pues la significancia estadística se ve afectada por el tamaño de la muestra y es más probable que un efecto se considere estadísticamente significativo cuando el tamaño de la muestra es

mayor, aunque la magnitud de la diferencia no sea realmente importante o significativa (Ellis, 2010).

Los grandes volúmenes de información en el ámbito de la investigación, que son posible gracias al avance de las tecnologías, requieren el uso del concepto de tamaño del efecto, debido a que el aumento excesivo del tamaño de muestra causa un aumento de la potencia, ocasionando un error de tipo 1 y, en consecuencia, conclusiones e interpretaciones erróneas de la información analizada. Por lo anterior, el objetivo de este ensayo es proponer el uso de simulaciones sobre una gran base de datos para la enseñanza y concientización de los estudiantes sobre otros conceptos que son necesarios de considerar al momento de decidir si se rechaza o no una hipótesis nula. Estos elementos debieran ser utilizados sin importar el tamaño de la base de datos analizada para dar sustento a los hallazgos encontrados en las investigaciones.

## **2. Enseñanza de las pruebas de hipótesis**

A continuación, se muestran algunas aportaciones referidas en la literatura de Didáctica de las Matemáticas, relacionadas con la prueba de hipótesis, sobre la manera en que se enseña y cómo se encuentra en los libros de texto, que se suelen usar como material de apoyo para los cursos de estadística.

### **2.1 Diferencias conceptuales de las pruebas de hipótesis**

La Didáctica de las Matemáticas ya tiene focalizada a la inferencia estadística mediante las pruebas de hipótesis como un punto que causa conflicto en la enseñanza-aprendizaje en los estudiantes. Como refiere Batanero (2001), no se debe a una enseñanza insuficiente del tema, ya que este problema también se encuentra en personas con una fuerte preparación estadística. El mismo autor menciona que el conflicto surge por la conjunción de dos métodos: el de Fisher y el de Neyman-Pearson.

En el mismo sentido, otras investigaciones (Albert et al., 2014; Ellis, 2010) refieren que las dificultades conceptuales en las pruebas de hipótesis están vinculadas a la filosofía y psicología, que la hacen susceptible de interpretaciones incorrectas. Además, su origen se debe a que es una mezcla informal de los contrastes de significación formales de Fisher y la teoría de Neyman-Pearson, así como conceptos e interpretaciones que no son parte de esta última. Todo lo anterior da lugar a los valores  $p$ , que hoy se usan para hacer inferencias estadísticas. Aunque ambos métodos se ubican dentro de la estadística clásica o frecuencial, las diferencias están en su razonamiento subyacente.

Al respecto, Goodman (2014) declara que el valor  $p$  se ideó como medida inferencial flexible, mientras que las pruebas de hipótesis se idearon como reglas de conducta y no de inferencia. Por eso, la conjunción de ambos métodos llevó a reinterpretar el valor  $p$  como una tasa observada de error y como medida de carácter probatorio, lo que ha impedido entender las consecuencias filosóficas de estos métodos, que son de uso generalizado. En la misma línea de observación, Romero y León (2018) destacan que es erróneo considerar que los valores de  $p$  puedan ser, a la vez, una tasa de error de falsos positivos (Neyman-Pearson) y una medida de la evidencia contra la hipótesis nula (Fisher).

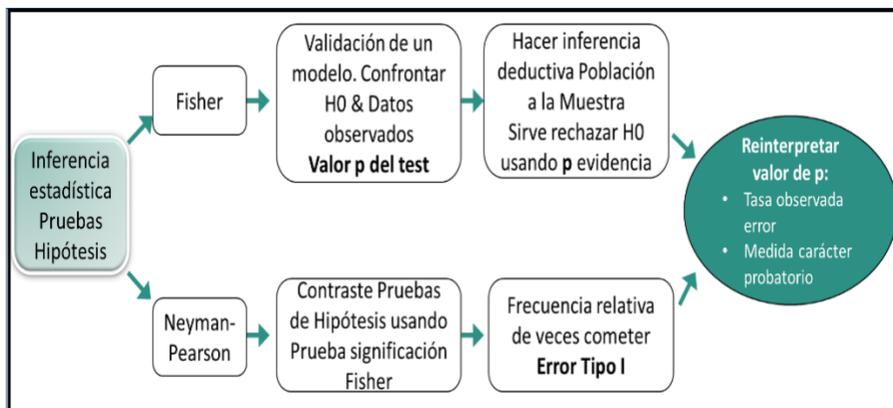
Como refiere la investigación de Albert et al., (2014), las pruebas de significación de Fisher se pueden considerar como procedimiento de validación de un modelo, donde dada su distribución y planteada la hipótesis nula, se examina si los datos son adecuados a ese modelo planteado. Es decir, confrontar la hipótesis nula con los datos observados, calculando la probabilidad de obtener un dato tan o más raro que el obtenido, dicho valor se le denominó el valor  $p$  de la prueba. En este contexto, el nivel de significación  $\alpha$  es una regla de decisión por la que se rechaza o no la hipótesis nula, y cuando  $p < \alpha$  se dice que los resultados son estadísticamente significativos.

En el enfoque propuesto por Fisher en 1925, las pruebas estadísticas son para hacer una inferencia deductiva de la población de las posibles muestras, a la muestra particular obtenida en cada caso. Su prueba sirve para rechazar una hipótesis nula, nunca para confirmarla, usando el valor  $p$  de la prueba como fuerza de evidencia en su contra.

Por otro lado, el enfoque de Neyman-Pearson hizo su versión contraste de pruebas de hipótesis usando la idea de Fisher de prueba de significación, pudiendo definir un  $\alpha$  a priori, para indicar la probabilidad de rechazar equivocadamente la hipótesis nula (error de tipo I) y la regla para tomar decisiones. Dicho método, entonces, requiere de una hipótesis alternativa. Existen dos tipos de errores que se pueden cometer tratando de encontrar la prueba con alto valor de potencia ( $1-\alpha$ ): tipo I y tipo II. Un resultado significativo implica, para Fisher, que los datos proporcionan evidencia en contra de la hipótesis nula, mientras que para Neyman-Pearson, establece la frecuencia relativa de veces que rechazaríamos una hipótesis nula verdadera cometiendo así un error de tipo I. La Figura 1 ilustra el propósito para el cual se crearon ambos métodos.

**Figura 1**

*Métodos de prueba de hipótesis según Fisher y Neyman-Pearson*



Como señalan Monterrey Gutiérrez et al., (2018), al ser la prueba de hipótesis el elemento de análisis de datos más utilizado, se ha producido pérdida de rigor y abuso en su uso. Además, muchas veces quienes lo usan desconocen sus fundamentos, lo que conduce a

obtener conclusiones absurdas, ilógicas e inconsistentes. El resultado final pone en entredicho la pertinencia y validez de la técnica en sí misma.

Las pruebas de significación estadística se utilizan para gestionar el riesgo de confundir la variación del muestreo aleatorio con efectos genuinos y el poder de dichas pruebas es afectado por varios parámetros, de los cuales el tamaño del efecto es solo uno. Por eso, sus resultados no pueden ser utilizados para informar conclusiones acerca de las magnitudes del efecto (Ellis, 2010).

La función de las pruebas, a decir de Monterrey Gutiérrez et al., (2018), es eliminar en los datos el efecto del azar y ver si aún la diferencia es digna de tomarse en cuenta. Por eso, se dice que es significativa. Un valor  $p$  pequeño indica que el azar es una explicación poco probable de las discrepancias encontradas y por eso se rechaza  $H_0$ . Eso funciona bien con muestras grandes (de aquí nuestro interés de *atacar* mediante el tamaño del efecto el rechazo de la  $H_0$  en el análisis de las grandes bases de datos). Sin embargo, en muestras pequeñas los niveles de información de la muestra son muy inestables porque el azar tiene un peso muy importante y los errores de estimación de los parámetros son mayores, mientras menor sea el tamaño de la muestra.

No sólo es la conjunción de los métodos de Fisher y Neyman-Pearson la causa de conflicto, también lo es que la atención se suele centrar en los contrastes de hipótesis, apoyándose únicamente en la significación estadística, sin estimar la magnitud de los efectos encontrada ni reflejando su importancia. En el mismo sentido, las investigaciones de Rüyü & Kula (2019), refieren que la persistencia de dichas omisiones en su cálculo y reporte se debe a la inercia, confusión conceptual, falta de otros instrumentos alternativos, así como por la inadecuada generalización del razonamiento lógico deductivo al razonamiento en la inferencia bajo incertidumbre.

A pesar de sus muchas limitaciones, las pruebas de significancia persisten porque proporcionan una base para comprobar que los resultados obtenidos a partir de muestras no se deben al azar y sugieren no abandonar el contraste de hipótesis, sino complementarlo proporcionando e interpretando el tamaño del efecto, publicar los valores  $p$  exactos, los intervalos de confianza y evaluando la replicabilidad de los resultados. Sin embargo, no ha tenido el impacto deseado, ya que los investigadores siguen sin incorporarlos en los artículos que mandan a revisión para publicar (Ellis, 2010).

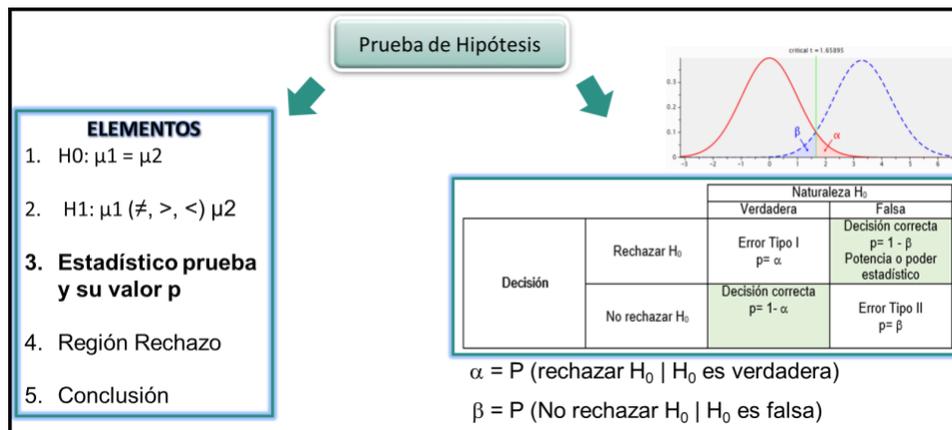
## **2.2 Concepto de prueba de hipótesis en los libros de estadística**

En relación con la enseñanza de la prueba de hipótesis, en los libros de estadística no se hace ninguna reseña respecto a cómo surge dicho método. Tampoco se hacen referencia a la diferencia entre ambos métodos, ni se mencionan las motivaciones que provocaron su surgimiento. Simplemente los conjuntan en uno solo y, de esa forma, se utiliza por investigadores en sus prácticas y por los educadores en su labor docente. Dichos métodos de enseñanza funcionan bien cuando los tamaños de muestra no son grandes, como los requeridos para hacer un análisis en una gran base de datos. Sin embargo, la enseñanza del tamaño del efecto da un mayor significado e importancia a los hallazgos encontrados en los resultados.

En el libro de Sullivan (2016) se refiere que investigaciones de todas las áreas hacen uso de las pruebas de hipótesis para analizar los datos obtenidos con el fin contrastar declaraciones respecto a una característica o parámetro de una o más poblaciones. Esto se hace basándose en muestras y probabilidad, por lo que no se tiene el cien por ciento de certeza que la afirmación es verdadera. Tan solo se puede determinar si los datos apoyan la declaración sobre el valor de un parámetro de la población o no.

En otros libros, como en el de Mendenhall et al., (2010), se explica el concepto de prueba de hipótesis de la siguiente manera: “tomar una decisión acerca del valor de un parámetro<sup>1</sup>, con base en alguna idea preconcebida acerca de cuál podría ser su valor”(p. 298). Junto con la explicación de los cinco elementos que la componen, y otros más, como en el libro de Rincón (2019), como una regla para decidir si aceptamos la hipótesis nula ( $H_0$ ) o la rechazamos en favor de la hipótesis alternativa ( $H_1$ ) pudiendo cometer dos tipos de errores: rechazar la  $H_0$  cuando es verdadera, denominado error tipo I ( $\alpha$  igual al tamaño de la región crítica o de rechazo y a esa probabilidad se le llama nivel de significancia) o no rechazar la hipótesis nula cuando es falsa, denominado error tipo II ( $\beta$ ), ambos métodos se muestran en la Figura 2.

**Figura 2**  
Prueba de hipótesis en los libros de texto



Nota. Adaptado de Mendenhall, Beaver & Beaver (2010); Rincón (2019).

<sup>1</sup> Medidas numéricas descriptivas que caracterizan a una población, como la media poblacional  $\mu$ , la desviación estándar poblacional  $\sigma$ .

## 2.3 Enseñanza de la prueba de hipótesis en el contexto de grandes bases de datos

En los primeros cursos de estadística se aprende que, si el tamaño de la muestra ( $n$ ) es grande, siempre se va a rechazar la hipótesis nula ( $H_0$ ). También se aprende que el estimador máximo verosímil es asintóticamente normal con media el verdadero valor del parámetro y varianza la inversa de la matriz de Fisher. Esta propiedad se usa para hacer pruebas de hipótesis. Sin embargo, al aumentar el tamaño de muestra en exceso, se llega a tener una potencia excesiva que lleva al extremo de rechazar  $H_0$  cuando es verdadera desde el punto de vista práctico (Gutiérrez Pulido & De la Vara Salazar, 2009; Monterrey Gutiérrez et al., 2018).

En el contexto del análisis de grandes bases de datos (cientos de miles de registros que se obtienen principalmente en áreas de las ciencias sociales, investigación de mercados y la educación) se convierten en un verdadero problema, cuando lo que se quiere es probar que, en realidad, no hay un efecto o cambio que nos lleve a rechazar la hipótesis nula cuando es verdadera. Esto se ha atacado definiendo otros conceptos, como el tamaño del efecto, donde se trata de ver que, aunque significativo, un efecto puede ser no importante. Como señala Ellis (2010), para evaluar la importancia práctica de un resultado, no es suficiente que conozcamos el tamaño del efecto, las magnitudes del efecto deben interpretarse para extraer significado.

Se desconoce la existencia de algún software o simulador que localice el punto en que se rechaza la hipótesis nula al incrementarse excesivamente el tamaño de muestra. Por eso, es relevante construir una herramienta que, con base en simulaciones, permita analizar el comportamiento del nivel de significancia, potencia de la prueba y tamaño del efecto para diferentes estructuras de un modelo estadístico y diferentes tamaños de muestra. Se podrían enseñar esos conceptos utilizando el software creado para posteriormente evaluar si su uso mejoró o facilitó la comprensión de los mismos. Lo anterior ayudaría a dar una mejor

interpretación práctica de los resultados, tomar mejores decisiones en consecuencia y poder fundamentar o justificar el porqué de estas.

Lo que se intentaría con ese simulador, es usar el número de observaciones utilizadas a partir del cual se rechaza la hipótesis nula para causarle un paradigma al estudiante cuando intente concluir si rechaza o no la hipótesis nula, como se le ha enseñado, y comprenda cómo dichos elementos no le sirven al aumentar una observación más en el modelo pues, a partir de ahí, se rechaza la hipótesis nula al tomar en cuenta solamente el resultado del estadístico de prueba y el valor  $p$  que le acompaña. Se espera que, de esta manera, recuerde siempre la importancia de analizar otros elementos como el tamaño del efecto, intervalos de confianza, potencia y valores de  $p$  exactos para tomar una decisión fundamentada de si es prudente, o no, rechazar la hipótesis nula. Y, en base a eso, resaltar la importancia de los hallazgos encontrados.

## **2.4 Metodología para la enseñanza de la prueba de hipótesis en el contexto de grandes bases de datos**

La investigación que se podría desarrollar sería de carácter mixto con observación participante y tendría dos objetivos. Uno, a nivel construcción de un simulador; y otro, a nivel didáctico, que presente situaciones que problematicen la toma de decisiones en los resultados de la prueba de significancia.

La idea es construir una herramienta que, con base en simulaciones, permita analizar el comportamiento del nivel de significancia, potencia de la prueba y tamaño del efecto para

diferentes estructuras del modelo estadístico y diferentes tamaños de muestra, el cual sería creado con el recurso de software R utilizando RStudio2.

El marco teórico en el que sustentaría la investigación es la Aproximación Instrumental (TAI) de uso de herramientas tecnológicas (Artigue, 2002, 2011). Se usaría para analizar los efectos el uso del simulador en los aprendizajes de los estudiantes respecto a la prueba de significancia, donde el simulador sería una variable.

La TAI es un enfoque teórico utilizado en la investigación en educación matemática para la integración de la tecnología digital en la enseñanza y el aprendizaje. Se preocupa por los aspectos instrumentales de la actividad de uso de una herramienta tecnológica por parte de un sujeto en un contexto educativo.

La TAI es resultado de la concatenación de dos teorías, la Teoría Antropológica de lo Didáctico (TAD) de Chevalard, referida en Castela (2017) y la Ergonomía Cognitiva de Rabardel, referida en Laisney & Chatoney (2018), que es la relacionada con el software. Su propósito es dar legitimidad educativa a las prácticas emergentes del uso de una herramienta de tecnología digital por parte del alumno en un contexto escolar. Por una parte, la Ergonomía Cognitiva da relevancia a las cuestiones instrumentales de la actividad y a la modelación que estas herramientas hacen sobre los procesos del aprendizaje. En cambio, la TAD se enfoca en estudiar qué hace y cómo hace el estudiante para resolver un problema matemático mediante cuatro componentes: Tarea, Técnica, Tecnología y Teoría. Es decir, analiza la producción matemática del estudiante. La praxeología de la TAD permitiría ver, mediante la producción matemática de los estudiantes, el grado de comprensión que tienen

---

<sup>2</sup> R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL

<https://www.R-project.org/>.

los estudiantes antes y después del uso de la herramienta propuesta; es lo que le da la legitimidad educativa.

Las tareas o actividades que se propondrían a los estudiantes se harían desde el marco de la TAI, con el fin de analizar el proceso de enseñanza y aprendizaje de los conceptos asociados al rechazo de la hipótesis nula, el cuál será uno de los instrumentos de recolección de datos. Las tareas se diseñarían de acuerdo con los modelos que se obtienen de un modelo estadístico. Dichas tareas se pondrían en hojas de papel con la finalidad de recuperar toda la producción matemática de los estudiantes. Eso se haría para evaluar, antes y después, del punto de corte o tamaño de muestra donde, debido al aumento de potencia estadística, se empieza a rechazar la hipótesis nula en la base de datos de donde se obtienen los modelos, con la finalidad de ver cómo determinan si rechazan la hipótesis nula en el modelo estadístico obtenido al analizar una gran base de datos.

También, de forma cuantitativa, se haría un análisis de los resultados obtenidos con un pre y post-test como una manera de cuantificar si es eficaz el uso del simulador. Además, se haría uso de algún cuestionario con escalamiento Likert para obtener la percepción de los participantes de la investigación en cuanto a su uso y lo aprendido.

### **3. Conclusiones**

La finalidad de este proyecto sería determinar en qué medida el uso de la simulación en un modelo estadístico obtenido de una gran base de datos, favorece la comprensión de los conceptos asociados al rechazo de la hipótesis nula en los estudiantes y responder así la pregunta de investigación.

Investigaciones como la de Borovcnik (2020), que han utilizado simulaciones para la enseñanza de conceptos estadísticos, han llegado a la conclusión de que la simulación puede corregir intuiciones de conceptos teóricos básicos inadecuadas, como es la probabilidad. Con las simulaciones se motiva a los estudiantes a resolver problemas

mediante el diseño de tareas y rutas de aprendizaje esperadas, que están centradas en el propósito del concepto que se quiere enseñar y, con ello, se convence a los estudiantes que los conceptos son útiles y les ayudan a resolver problemas. Todo lo anterior les da un sentido, permitiéndoles completar su red cognitiva para ese concepto y obtener una comprensión más amplia, que les permitiría sacar sus conclusiones.

Por su parte, Batanero (2018) refiere que el uso de la simulación para comenzar la enseñanza de la prueba de hipótesis como una aproximación informal, ayuda al alumno con conocimientos básicos previos de estadística, al entendimiento de la lógica del proceso y los conceptos asociados a ésta y a la adquisición de una comprensión razonable de dicho concepto. También hace hincapié en señalar que una distribución muestral, obtenida mediante simulación, es diferente a la teórica, por lo que se añade un error de estimación ocasionado por la fluctuación aleatoria del muestreo. La distribución muestral empírica obtenida por simulación converge a la teórica cuando el número de muestras tiende al infinito.

Se espera que este proyecto aporte su granito de arena para generar un cambio en la forma en la que se enseña y aparece en los libros de estadística la enseñanza de la prueba de hipótesis. De manera que, si se pone énfasis en la enseñanza del tamaño del efecto como medida alternativa y complementaria, en lugar del valor  $p$ , se obtendría una mejor decisión al rechazar la hipótesis nula de un estudio y, por ende, mejores conclusiones e interpretación de los hallazgos encontrados.

En consecuencia, se lograría la cultura de utilizar y reportar dicho índice a manera de justificación de las conclusiones e importancia de los hallazgos encontrados en un estudio, evitando cometer errores de tipo I, al aumentar el tamaño de la potencia debido al aumento del tamaño de muestra en grandes bases de datos.

## Referencias

- Albert, J. A., Rodríguez, M. I., Hernández, S. A., Torres, J. L., Alvarado, H., Albrecht, G., & Ruiz, A. M. (2014). Red de colaboración en investigación de la educación estadística. *Acta Latinoamericana de Matemática Educativa*, 0(0), 389–398. <http://funes.uniandes.edu.co/5340/1/AlbertRedALME2014.pdf>
- Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, 7(245), 1–26. <https://doi.org/10.1023/A:1022103903080>
- Artigue, M. (2011). Tecnología y enseñanza de las matemáticas: desarrollo. *Cuadernos de Investigación y Formación en Educación Matemática*, 6(8), 13–33. <https://doi.org/10.1023/A:1022103903080>
- Batanero, C. (2001). *Didáctica de la Estadística*. Universidad de Granada. <https://doi.org/10.13140/RG.2.1.3946.7044>
- Batanero, C. (2018). Treinta años de investigación didáctica sobre el análisis inferencial de los datos. *Sociedad Mexicana de Investigación y Divulgación de la Educación Matemática*, 0(0), 196–209.
- Borovcnik, M. (2020). Contexts, analogies, and tasks that expose the purpose of the key concepts of probability. *Zetetiké*, 0(0), 1–24. <https://doi.org/10.20396/zet.v28i0.8657607>
- Castela, C. (2017). La Teoría Antropológica de lo Didáctico: Herramientas para las ciencias de la educación. *Herramientas para las Ciencias de la Educación*, 59(0), 8–15.
- Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press. [www.cambridge.org/9780521142465](http://www.cambridge.org/9780521142465)
- Goodman, S. . (2014). Valores P, pruebas de hipótesis y verosimilitud: las consecuencias para la epidemiología de un debate histórico ignorado. *Boletín Oficina Sanitaria Panamericana*, 118(2),

141–155. <https://doi.org/10.1002/9781118445112.stat06809>

- Gutiérrez Pulido, H., & De la Vara Salazar, R. (2009). *Control estadístico de calidad y seis sigma* (2ª ed.). McGraw Hill.
- Laisney, P., & Chatoney, M. (2018). Instrumented activity and theory of instrument of Pierre Rabardel. In *Philosophy of technology for technology education to Sense/Brill* (Vol. 0, Issue 0). <https://hal-amu.archives-ouvertes.fr/hal-01903109>
- Leenen, I. (2012). La prueba de la hipótesis nula y sus alternativas: revisión de algunas críticas y su relevancia para las ciencias médicas. Investigación en educación médica. *Investigación en Educación Médica*, 1(4), 225–234.
- Mendenhall, W., Beaver, E. J., & Beaver, B. M. (2010). *Introducción a la probabilidad y estadística* (CENGAGE Learning (ed.); 13a ed.). <http://latinoamerica.cengage.com>
- Monterrey Gutiérrez, P. A., Cortés Sanabria, L. Y., & Díaz Hernández, M. E. (2018). Utilidad y limitaciones de las pruebas de hipótesis en la epidemiología nutricional ¿Cómo proceder frente a un problema? *Perspectivas en Nutrición Humana, Universidad de Antioquia*, 9, 72–87. <https://revistas.udea.edu.co/index.php/nutricion/article/view/336392>
- Rincón, L. (2019). *Una introducción a la estadística inferencial*. Facultad de Ciencias (1ª ed.) UNAM.
- Romero, M. J., & León, L. E. (2018). Algorithms in the handling of samples and variables in biostatistic. *Revista 16 de Abril*, 57(269), 177–194. <https://www.medigraphic.com/pdfs/abril/abr-2018/abr18269h.pdf>
- Rüya, K., & Kula, F. (2019). Why does statistical inference remain difficult? A textbook analysis for the phases of inference. *Actas Del XI Congreso de la Sociedad Europea de Investigación en Educación Matemática (CERME11), TWG05(22)*. <https://hal.archives-ouvertes.fr/hal-02412780>
- Sullivan, M. I. (2016). *Statistics informed decisions using data* (5a. ed.) Pearson.