

# El valor $p$ : ¿cómo analizarlo para separarse del positivismo extremo e inductivismo ingenuo?

Fernando Padilla-Santamaría<sup>a,\*</sup>

Facultad de Medicina



## Resumen

A pesar de las nuevas reflexiones y corrientes ideológicas en las ciencias médicas, muchos investigadores continúan apegados firmemente al positivismo científico, al inductivismo ingenuo y a miradas totalmente reduccionistas, donde el uso de estadística inferencial se considera casi indispensable para juzgar una investigación “de calidad”, y dentro de ella, estimar al famoso valor  $p$  como el número que determina que un estudio es “bueno” o “malo”, que vale la pena o no, que salió “bien” o “mal”. Aunque la estadística inferencial es la más frecuente en las ciencias médicas, muchos investigadores continúan con problemas epistemológicos para la interpretación del valor  $p$  y la toma de decisiones estadísticas; por ello, el objetivo principal de este trabajo es brindar una reflexión y análisis dinámico de qué es, cómo se obtiene, cómo suele interpretarse y cómo debería interpretarse el valor  $p$ .

Cabe destacar que este escrito no pretende enseñar estadística, sino que intenta cambiar la forma en que los estudiantes y profesionales de la salud interpretan la estadística inferencial, con el fin de fomentar la lectura crítica y de esta forma dar armas para el aprendizaje autodidacta. Para llegar al adecuado análisis del valor  $p$ , a lo largo del trabajo realicé un repaso general y gráfico acerca de la construcción de hipótesis, la distribución normal y las pruebas de hipótesis.

Si bien por el simple hecho de que este trabajo hable de estadística inferencial ya lo convierte (hasta cierto punto) en un artículo positivista, deseo que la nueva enseñanza en esta área permita formar nuevos profesionales e investigadores con visiones más amplias de la investigación, y así terminar con el fomento del reduccionismo e inductivismo ingenuo.

<sup>a</sup> Departamento de Atención a la Salud, División de Ciencias Biológicas y de la Salud, Universidad Autónoma Metropolitana Unidad Xochimilco. Ciudad de México, México.

ORCID ID:

† <https://orcid.org/0000-0001-5535-1332>

Recibido: 2-abril-2022. Aceptado: 11-agosto-2022.

\* Autor de correspondencia: Fernando Padilla-Santamaría.

Dirección postal: Calzada del Hueso 1100, Col. Villa Quietud, Alcaldía Coyoacán, C.P. 04960. Ciudad de México, México.

Teléfono (móvil): (+52 1) 55 4536 7901.

Correo electrónico: fernando.psantamaria23@gmail.com

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Palabras clave:** Pruebas de hipótesis; intervalos de confianza; distribución normal; metodologías cuantitativas.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## The $p$ -value: how to analyze it to separate from extreme positivism and naive inductivism?

### Abstract

Despite the new reflections and ideological currents in medical sciences, many researchers remain firmly attached to scientific positivism, naive inductivism and totally reductionist views, where the use of inferential statistics is considered almost indispensable to consider a “high quality” research, and within it, consider the famous  $p$ -value as the number that determines that a study is “good” or “bad”, that it is worthwhile or not, that it came out “good” or “bad”. Although inferential statistics is the most frequent in medical sciences, many researchers continue with epistemological problems for the  $p$ -value interpretation and make statistical decisions; Therefore, the main objective of this work is to provide a reflection and dynamic

analysis of what is  $p$ -value, how it is obtained, how it is usually interpreted, and how it should be interpreted.

It should be noted that this paper does not intend to teach statistics, but rather tries to change the way in which students and health professionals interpret inferential statistics, in order to encourage critical reading and thus provide weapons for self-taught learning. To arrive at the appropriate analysis of the  $p$ -value, throughout the work I carry out a general and graphic review about the construction of hypotheses, the normal distribution and the hypothesis tests.

Although the simple fact that this work talks about inferential statistics already makes it (until a certain point) a positivist article, I hope that the new teaching in this area will allow the training of new professionals and researchers with broader visions of research, and thus, ending the promotion of reductionism and naive inductivism.

**Keywords:** Hypothesis tests; confidence intervals; normal distribution; quantitative methodologies.

This is an Open Access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## INTRODUCCIÓN

La búsqueda de la verdad ha sido la tarea principal de las ciencias, donde la medición (metodologías cuantitativas) fue por mucho tiempo considerada como el único sendero certero para caminar. No obstante, actualmente sabemos que *verdad* es un término delicado en ciencias<sup>1</sup>, además de que el *inductivismo ingenuo* criticado por Alan Chalmers ha dejado claro que el primer error de cualquier investigador es considerar la observación (y por ende, la medición) como el inicio del estudio de un fenómeno, pues la investigación realmente parte de la teoría: no es posible buscar algo que ni siquiera tengamos idea que pueda existir<sup>2</sup>.

Gran parte de los errores que competen a muchos investigadores es apegarse firmemente al *positivismo científico*. Esta corriente de pensamiento surgió como doctrina con Auguste Comte en el siglo XIX

describiendo los estados teóricos de la humanidad, donde el último de ellos es el estado científico, así, describió la física social (ahora sociología): para Comte, los métodos no bastan para dar bases a las ciencias, sino que deben ir acompañados de las doctrinas, considerándolos inseparables; por ello, planteó que para el estudio de los fenómenos no basta con su historia individual, sino también se debe incorporar la historia de las sociedades<sup>3</sup>. La característica principal del positivismo científico es considerar como verdad todo aquello que puede analizarse mediante la medición, la lógica y razón (es decir, todo aquello que podemos observar, tocar, oler, degustar y escuchar); por ello, dicho movimiento considera la metodología cuantitativa como la única real fuente de conocimiento<sup>1-3</sup> y, dentro de ella, toma al famoso valor  $p$  como el número que califica un estudio como “bueno” o “malo”, que vale o no

vale la pena, que salió “bien” o “mal”, e incluso como el único con el “poder” de determinar *descubrimientos*. Sabemos que la estadística inferencial es la más frecuentemente empleada en las ciencias médicas; a pesar de ello, muchos investigadores continúan con problemas epistemológicos para la interpretación del valor  $p$  y la toma de decisiones estadísticas, pues el pensamiento reduccionista y la falta de razonamiento hace que, lejos de analizar los resultados, se tomen decisiones absolutistas, provocando la distorsión de la naturaleza del fenómeno estudiado<sup>4</sup>. Por ello, para fines de este trabajo llamaré *positivistas extremos* a los profesionales que ven a la estadística inferencial de forma reduccionista y absolutista.

Con el fin de intentar cambiar dicha perspectiva de la estadística inferencial en estudiantes, docentes e investigadores de las ciencias médicas, los objetivos del presente escrito son: 1) exponer –a *grosso modo* y de forma gráfica– qué es y cómo se obtiene el valor  $p$ , retomando cómo se construyen las hipótesis, qué es la distribución normal y en qué consisten las pruebas de hipótesis; 2) cómo suele interpretarse de forma cotidiana por positivistas extremos; 3) las falsas creencias que se desarrollan en torno a dicho número,  $y$ ; 4) brindar al lector información pertinente con el fin de ampliar la visión de la interpretación del valor  $p$ , y así separarse de la mirada reduccionista.

Cabe destacar que este artículo no pretende enseñar estadística, únicamente intenta dejar con mayor claridad qué es y cómo se sugiere ver e interpretar dicho valor, con el objetivo de mejorar la lectura crítica y reflexiva de artículos de investigación, así como modificar la forma de mirar la estadística inferencial en el desarrollo de protocolos de investigación y tesis de grado y/o posgrado.

### Recordando la construcción de hipótesis

Antes de discutir propiamente el valor  $p$ , es necesario recordar que para el análisis estadístico se deben construir 2 hipótesis: la nula o estadística ( $H_0$ ), y la alternativa o de investigación ( $H_1$ ). En estadística inferencial, las decisiones se toman en torno a la  $H_0$ , pues lo que se busca es decidir si es posible rechazarla o no. La  $H_0$  siempre plantea igualdad, mientras que la  $H_1$  apuesta por diferencias<sup>5</sup> (tabla 1).

Existe un error común al momento de interpretar la decisión estadística sobre la  $H_0$ , pues suele decirse

**Tabla 1.** Comparación de las diferentes posibilidades de hipótesis al comparar la variable X con la variable Y

$H_0$	$H_1$
$X = Y$	$X \neq Y$
$X \leq Y$	$X > Y$
$X \geq Y$	$X < Y$

que la  $H_0$  “se acepta” o “se rechaza”. Recordemos que la estadística inferencial plantea probabilidades; el hecho de “aceptar” una  $H_0$  significa tomarla junto con sus posibles errores estadísticos y metodológicos, por lo que *aceptar* sería un término inadecuado. Por ello, con apoyo de un ejemplo simple y superficial, deseo explicar la sugerencia de enunciados ideales para expresar la decisión estadística sobre la  $H_0$ :

Un grupo de investigadores realizó un ensayo clínico controlado donde probaron el fármaco A-203 en pacientes con diabetes mellitus tipo 2; el objetivo fue demostrar que dicho medicamento controlaba mejor los niveles de glucosa sérica que la pioglitazona y un placebo. Se obtuvieron los siguientes resultados: el fármaco A-203 tuvo un efecto similar a la pioglitazona, y tanto el A-203 como la pioglitazona presentaron efectos benéficos sobre la glucemia, comparados con el placebo. Para este ejemplo, no deseo colocar aún valores  $p$  ni intervalos de confianza, ya que los abordaré más adelante en este escrito con mejor claridad; con la información aquí brindada, la interpretación de la decisión hipotética en términos estadísticos fue la siguiente:

- Para la comparación entre pioglitazona y A-203: **No se rechaza** la  $H_0$  a un nivel de confianza del X%. → El efecto hipoglucemiante de A-203 es igual al observado con pioglitazona (al menos al X% de confianza). Suponiendo que el intervalo de confianza fuera del 95%, esta interpretación indicaría que, si le diéramos a 100 personas A-203, al menos 95 de ellas tendrían el mismo efecto hipoglucemiante que la pioglitazona.
- Para la comparación entre pioglitazona/A-203 y el placebo: **Se rechaza** la  $H_0$  a un nivel de confianza del X%. → La pioglitazona y el A-203 tienen un efecto hipoglucemiante significativamente mayor al placebo (al menos al X% de confianza). Al igual que en la interpretación anterior,

suponiendo que el intervalo de confianza fuera del 95%, si a 100 personas les diéramos el placebo con fines hipoglucemiantes, en menos de 5 individuos posiblemente se observe un efecto benéfico, mientras que en 95 o más funcionará la pioglitazona y el A-203.

Es necesario aclarar que para este ejemplo usé un placebo que posiblemente funcione en una pequeña minoría como agente hipoglucemiante; sin embargo, el lector debe saber que en la realidad un placebo (agua con sal o azúcar, como los más comunes) no tendrían efecto hipoglucemiante alguno. Más adelante en este escrito, usted podrá comprender mejor los intervalos de confianza, por lo que al finalizar la lectura, recomiendo releer este ejemplo.

### ¿Qué es y de dónde sale el valor $p$ ?

El valor  $p$  (término propuesto por Fisher<sup>6</sup>) es un número que indica la probabilidad de estar equivocado al rechazar la  $H_0$ : entre más pequeño sea el valor  $p$  (más cercano a 0) la probabilidad de estar equivocado al rechazar la  $H_0$  es menor, es decir, que la probabilidad de estar equivocado al decir que la variable  $X$  se relaciona con la variable  $Y$  es poca. El valor  $p$  se obtiene a partir de las pruebas de hipótesis, por lo que también será necesario recordar la distribución normal y la no normal.

El modelo gráfico de la función matemática descrita por Carl Friedrich Gauss (mejor conocido como “campana de Gauss”) expone la distribución normal (**figura 1A**), donde la línea que divide la campana exactamente a la mitad corresponde a las medidas de tendencia central (es decir, todas son teóricamente iguales)<sup>5,7</sup>. En el caso de aquellas variables que no siguen una distribución normal, las medidas de tendencia central no son teóricamente iguales<sup>8</sup>, tal como se observa en la **figura 1B**. Este es precisamente el fundamento para decidir cuál medida de tendencia central se debe tomar para analizar una variable cuantitativa: se utilizará la media si la variable de interés sigue una distribución normal, mientras que se utilizará la mediana si esta no sigue una distribución normal; ¿por qué? Observe nuevamente en la **figura 1** cuál es la medida de tendencia central que parte a la mitad cada distribución.

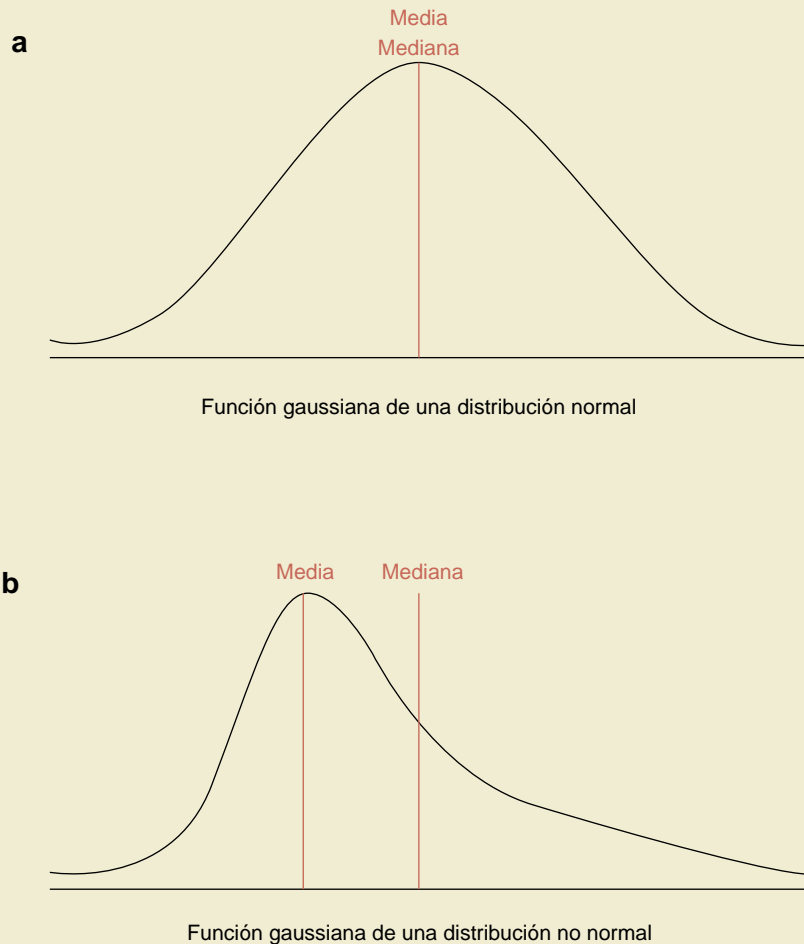
Ya que recordamos brevemente la constitución

de la función gaussiana, ahora es necesario aclarar cómo se compone el área bajo la curva cuando se trabaja a diferentes niveles o intervalos de confianza. Los intervalos de confianza representan la variabilidad entre el valor obtenido y el posible valor real de una variable en una población específica<sup>9</sup>. En las ciencias biológicas y de la salud, se ha acordado de forma internacional que el nivel de confianza más bajo para determinar que las diferencias observadas son significativas es al 95%<sup>5,9</sup>.

Es pertinente mencionar que existen ocasiones excepcionales donde el uso de estadística inferencial no es necesario para demostrar diferencias, por ejemplo: se desea buscar si existen diferencias entre la estatura de 50 personas pigmeas que miden todos exactamente 1.30 m comparadas con 50 personas asiáticas que miden todos exactamente 1.70 m. En este ejemplo es claro que en los datos no existe variabilidad debido a que todos son exactamente iguales en cada grupo, además de que cada uno tiene el mismo número de individuos; para este escenario, no vale la pena usar estadística inferencial (valores  $p$  ni intervalos de confianza), ya que es claro que las personas pigmeas son más bajas de estatura en comparación con las personas asiáticas, teniendo de diferencia exactamente 40 cm. Por su parte, en caso de que se busquen diferencias de estatura en 45 personas pigmeas que miden entre 1.35-1.60 m VS 52 personas asiáticas que miden 1.50-1.75 m, dada la variabilidad de los datos, en este caso sí es considerado pertinente emplear estadística inferencial, ya que es necesario buscar qué tanto se parecen ambos grupos de mediciones y si es posible considerarlas diferentes o similares según los diferentes niveles de confianza que queramos usar.

En la **figura 2** se puede observar la composición del área bajo la curva cuando se trabaja a niveles de confianza del 95%, 99% y 99.99%; este último valor corresponde al máximo nivel de confianza. Como se comentó anteriormente, en bioestadística se trabaja con probabilidades, por lo que es un error filosófico y sistemático considerar un nivel de confianza del 100%. El área bajo la curva siempre será igual a 1, por lo que un nivel de confianza del 95% será equivalente a 0.95 del área bajo la curva, un nivel de confianza del 99% será equivalente a 0.99, y el máximo nivel de confianza (99.99%) será equivalente a 0.9999.

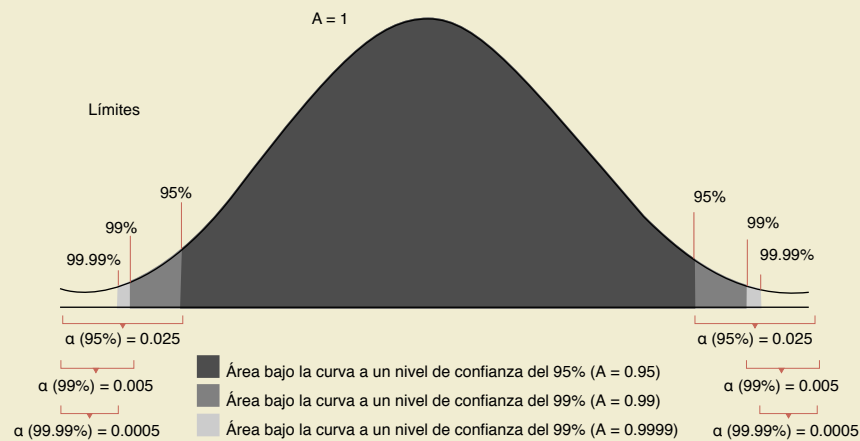
**Figura 1.** Ilustraciones de la función matemática de Gauss correspondientes a una variable cuyos datos siguen una distribución normal (a) y aquellos que no siguen una distribución normal (b). Nótese la posición diferente de las medidas de tendencia central en cada tipo de distribución



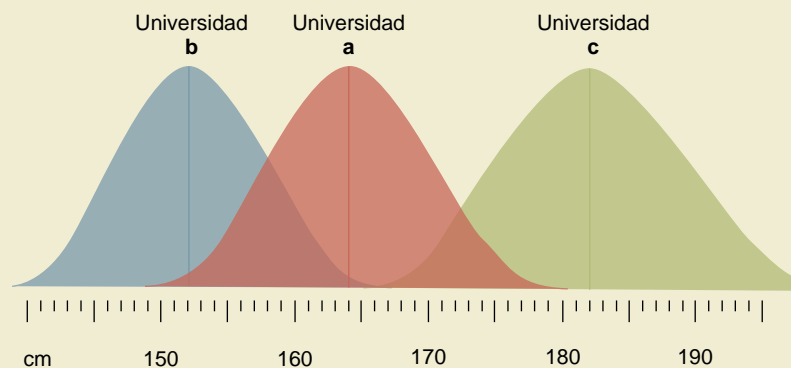
Partiendo de lo anterior, ¿dónde queda el valor restante para sumar 1 en el área bajo la curva? ¡Aquí comienza la comprensión del valor  $p$ ! Como se observa en la **figura 2**, los valores  $\alpha$  se encuentran en las *colitas* de la campana y cada uno (en el caso del nivel de confianza al 95%) equivale a 0.025; así,  $0.95 + 0.025 + 0.025 = 1$ . No obstante, si se suma  $0.025 + 0.025 = 0.05$ , este número sería el valor  $p$  máximo para trabajar al 95% de confianza y decidir si la  $H_0$  se rechaza o no.

Al momento de comparar 2 o más grupos/variables, las curvas de Gauss se sobreponen; las pruebas de hipótesis consisten en medir qué tan diferente es una distribución de la otra, de ahí que existan pruebas paramétricas (para variables cuantitativas continuas con distribución normal) y no paramétricas (para variables cuantitativas discontinuas o para variables cuantitativas continuas sin distribución normal). De esta forma, el valor  $p$  será alto cuando las curvas compartan muchos datos, y será bajo

**Figura 2.** Ilustración de la función matemática de Gauss de una variable con distribución normal, los límites de intervalos de confianza al 95%, 99% y 99.99%, así como los valores  $\alpha$  correspondientes a cada nivel de confianza



**Figura 3.** Ilustración de los hallazgos del estudio ejemplo desarrollado en las universidades **a**, **b** y **c**, donde se observa la distribución de tallas encontradas en cada población



cuanto las curvas compartan pocos datos (es decir, que los datos compartidos abarquen solo las *colitas*); en un ejemplo que se encuentra en el siguiente apartado se ilustrará mejor esta explicación. Dado que profundizar en las pruebas de hipótesis paramétricas y no paramétricas provocaría salir de los objetivos de este escrito, recomiendo consultar el artículo realizado por Flores-Ruiz E, et al. (2017)<sup>10</sup>, donde se expone una tabla muy práctica que ayuda

a decidir qué prueba de hipótesis es la ideal para cada escenario.

Ahora podemos entender a qué se refieren los investigadores cuando se enuncia en los artículos: “se trabajó al 95% de confianza, por lo que los valores  $p < 0.05$  fueron considerados estadísticamente significativos”. De esta forma, si el valor  $p$  de la prueba de hipótesis es menor a 0.05, se puede rechazar la  $H_0$  al 95% de confianza; en caso contrario, la  $H_0$  no se re-



chaza y, por tanto, no existen diferencias estadísticamente significativas (al menos al 95% de confianza).

### Un ejemplo no real de estudio inferencial

Un grupo de investigadores realizó un estudio acerca de la talla de los estudiantes de la universidad A: se encontró que la media ( $\mu$ ) fue de 1.64 m y mediana (Me) de 1.6 m. Decidieron comparar la estatura de esta población con otras, por lo que se trasladaron a 2 universidades más: 1) la universidad B, donde se encontró  $\mu = 1.52$  m y  $Me = 1.45$  m; y 2) la universidad C,  $\mu = 1.82$  m y  $Me = 1.80$  m. Dados los hallazgos, la distribución de las mediciones de tallas en estudiantes de las 3 universidades se expresa en la **figura 3**.

Como es posible apreciar, las curvas de Gauss correspondientes a las universidades A y B son muy parecidas y comparten muchos datos (observe la zona donde convergen ambas curvas, es decir, donde se mezcla el color rojo con el azul); no obstante, si se compara la estatura de los estudiantes de la universidad A y B con la de los estudiantes de la universidad C, es posible notar que los datos similares son pocos y las curvas comparten únicamente las *colitas*, es decir, que para considerar ambas distribuciones diferentes, estas deben compartir cuando mucho el 5% de todos los valores. A partir de esta representación gráfica, usted ya puede inferir cuáles serán las diferencias estadísticamente significativas y cuáles no, al comparar las 3 poblaciones.

Debido a que la variable de interés fue cuantitativa continua y seguía una distribución normal, los investigadores decidieron aplicar la prueba *t* de student a un nivel de confianza del 95% para determinar si existieron diferencias estadísticamente significativas entre las 3 poblaciones. Se obtuvo lo siguiente: A VS B  $p = 0.648$ , A VS C  $p = 0.0541$ , y B VS C  $p = 0.004$ . De esta forma, los investigadores determinaron que las tallas encontradas entre los estudiantes de la universidad A y B son similares, mientras que en la universidad C son significativamente diferentes solamente respecto a la universidad B. Observe los valores *p* nuevamente y reflexione: ¿está de acuerdo con que los investigadores únicamente tomaran el valor  $p = 0.004$  como el único realmente “valioso”? La respuesta la comentamos más adelante en este escrito.

### Interpretación del valor *p* por positivistas extremos e inductivistas ingenuos

Como se comentó anteriormente, para el investigador positivista e inductivista ingenuo, la observación es el medio principal de la obtención del conocimiento y no permite que su horizonte vaya más allá de lo que es tangible y medible, sin tomar en cuenta aspectos cualitativos que pueden influir en el fenómeno estudiado.

Alan Chalmers cita un ejemplo clásico de cómo el positivista e inductivista ingenuo interpreta las observaciones: un pavo es llevado a una granja donde el lunes le dan de comer a las 9:00 am, el martes a las 9:00 am, el miércoles a las 9:00 am y el viernes a las 9:00 am; con base en lo anterior, ¿a qué hora piensa que le den de comer al pavo el sábado y el domingo? ¡Exacto! ¡A las 9:00 am! Por lo que el pavo sabe que al día siguiente llegará su comida a las 9:00 am por el resto de sus días. No obstante, llegó diciembre y el pavo es decapitado a las 9:00 am del día 24<sup>2</sup>.

Con el ejemplo anterior, se expone la visión de un investigador positivista e inductivista ingenuo, donde una serie de observaciones darán lugar a la “verdad absoluta” que puede repetirse en cualquier lugar y momento. La muerte del pavo representa la variante de dicha observación, pues cualquier pavo positivista e inductivista ingenuo hubiese apostado –al máximo nivel de confianza ante tantos días previos de observación– que ese día a las 9:00 am recibiría alimento, mas no que se convertiría en la cena navideña. Así, es necesario comprender que toda investigación es una aproximación de la realidad y los resultados deben analizarse e interpretarse con precaución, y nunca de forma absolutista.

Por ello, el filósofo Karl Popper introdujo el falsacionismo científico, donde en lugar de buscar lo absoluto, se busca la probabilidad y siempre se da lugar a posibles fallas, falsaciones, errores o variabilidad en los hechos; por ello, según Popper, entre más falsable sea un enunciado, es mejor<sup>11</sup>.

Teniendo como base lo anterior, usted ya podrá inferir la forma en la que un positivista e inductivista ingenuo ve e interpreta el valor *p*: “si el número no es menor a 0.05, no puedes decir que hay diferencias”, “no te salió tu estudio, ya que ninguno de tus valores *p* es menor a 0.05”, “deja ya la *paja* y mejor vamos a lo importante, ¿dónde están tus valores *p*?”, “¿por

qué no encontraste valores  $p$  menores a 0.05? todos los estudios publicados dicen lo contrario ¡tu tesis está mal!”, entre muchas otras expresiones.

### Interpretación ideal del valor $p$

Como ya se ha comentado, el valor  $p$  no es una cifra absoluta que marque un punto de decisión rotundo sobre la realidad; lo único que simboliza es un apoyo para decidir única y exclusivamente sobre la  $H_0$ .

Retomando el ejemplo del estudio donde se midieron estaturas de estudiantes de 3 universidades, la conclusión “contundente” fue que se observaron diferencias estadísticamente significativas únicamente al comparar la universidad B con la C ( $p = 0.004$ ), pero ¿qué pasa entonces con la  $p = 0.0541$  obtenida de la comparación entre las mediciones obtenidas en la universidad A y la C? ¿Quedó muy cerca de la “significancia estadística”? ¿O no? Ante ello, se habla de la *tendencia a la significancia estadística*<sup>12</sup>.

La tendencia a la significancia estadística no es más que otra forma del positivismo dentro de la investigación, pues se expresa la posibilidad de que, si el estudio es replicado con un mayor tamaño muestral, mejores criterios de selección y/o diferentes métodos de análisis, la relación se encuentre al nivel de confianza deseado; sin embargo, ¿todo es culpa de la muestra y las pruebas estadísticas? ¿El investigador –siendo humano mortal– no tiene errores? Dejaré estas preguntas abiertas al lector.

Además, si bien ahora podemos saber que la estatura de los estudiantes de la universidad B es significativamente diferente a la encontrada en los estudiantes de la universidad C, así como la estatura medida en la universidad A tiene tendencia a la significancia estadística respecto a lo encontrado en la universidad C, ¿hacia dónde se *inclina la balanza*? Decir que “la variable X se asoció a la variable Y” o “la variable X es diferente a la variable Y” ofrece la mitad de la información total, pues únicamente nos plantea diferencia, pero no nos dice dónde es mayor o de qué forma se relacionan ambas variables; por ello, la estadística descriptiva juega un papel crucial en la interpretación de los valores  $p$ . Observe nuevamente la **figura 3** y revise las mediciones de cada distribución; ahora, podemos decir que los estudiantes de la universidad C son significativamente más altos que en la universidad

B, mientras que los estudiantes de la universidad C tienden a ser más altos que los estudiantes de la universidad A.

Ahora bien, aunque los hallazgos del estudio ejemplo rebelan las diferencias de estaturas entre las 3 universidades, el valor  $p$  tampoco ofrece la explicación a dicho fenómeno, es decir, que no puede responder a la pregunta ¿por qué unos son más altos que otros? La causalidad no solo requiere veracidad del análisis de datos, sino también confianza en su estabilidad, o sea, que otros tipos de pruebas no la pongan en duda, criterios que el valor  $p$  por sí solo no puede cumplir. Para poder determinar adecuadamente causalidad y evitar conjeturas espurias se requiere tomar en cuenta los siguientes elementos: 1) precedencia temporal (la causa siempre precede al efecto), 2) asociación (la variación de la causa se relaciona con la variación del efecto), 3) aislamiento de la relación causa-efecto (no hay explicaciones alternativas plausibles, como los efectos de las variables de confusión) y 4) replicación (los 3 puntos anteriores deben ser replicables en cualquier otra muestra)<sup>13</sup>. De forma adicional a lo anterior, también se ha planteado que debe explicarse claramente el mecanismo porque se da la relación causa-efecto<sup>14</sup>.

A pesar de lo anterior, es necesario aclarar que en muchos casos (principalmente en investigaciones socio-médicas) tomar en cuenta los puntos anteriores para una adecuada asociación causal no es suficiente, ya que existen otras variables y categorías que deben incluirse al análisis para mejorar la visión del fenómeno estudiado; por ello, la investigación cualitativa funge un papel crucial.

En el caso del ejemplo de las tallas medidas en las 3 universidades, la investigación cualitativa permitiría estudiar el origen étnico, la situación territorial, la alimentación y las características físicas de los padres (por mencionar solo algunas) cuya información permitiría posiblemente dar explicación a lo encontrado cuantitativamente, y así, evitar patologizar las variantes posiblemente consideradas “anormales” (por ejemplo, se puede inferir errónea e irreverentemente que todos los estudiantes de la universidad C tienen síndrome de Marfan, lo que explicaría la estatura significativamente más alta).

Por otra parte, si bien el valor  $p$  plantea diferencias estadísticas significativas o no, es necesario que



usted reflexione que dicha relación –falsa o no– se limita a evaluar una variable X con una variable Y, es decir, que existe un “lazo” que une ambas variables, pero inicialmente no se ofrece más allá de esta información, por lo que (siguiendo la metáfora del lazo) preguntas como las siguientes no pueden ser respondidas por el valor *p*: ¿de qué material es ese lazo?, ¿cuántas hebras tiene?, ¿es liso o trenzado?, ¿hay hebras rotas?, ¿realmente existe ese lazo o es una relación falsa por variables confusas? Entre muchas otras. Por ello, los avances en el conocimiento tardan décadas para aproximarse lo mayor posible a la realidad.

Aunque en este escrito abordo únicamente la significancia estadística, es necesario recordar que no es lo único valioso existente en la investigación, ya que también hay resultados con significancia clínica o significancia educativa, se realice o no estadística inferencial. Un ejemplo real y actual para hablar de la significancia clínica es el estudio desarrollado por el grupo RECOVERY, donde estudiaron los efectos del uso de dexametasona en pacientes con cuadros leves (ambulatorios con y sin oxígeno suplementario) y graves (hospitalizados con ventilación mecánica invasiva) de COVID-19: se observó que, en pacientes con ventilación mecánica invasiva, la dexametasona disminuyó el riesgo de muerte (riesgo = 0.64, IC95% 0.51-0.81), mientras que en pacientes ambulatorios sin oxígeno suplementario no simbolizó protección ni riesgo de muerte (riesgo = 1.19, IC95% 0.92-1.55); no obstante, a pesar de no encontrar significancia estadística (riesgo o protección de la dexametasona en pacientes ambulatorios), sí existe mayor mortalidad si esta es administrada<sup>15</sup>; así, la significancia clínica que ofrecen estos resultados es decidir no usar dexametasona en pacientes con COVID-19 leve, a pesar de los resultados de estadística inferencial. Por su parte, respecto a la significancia educativa, imagine un estudio donde se ponen a prueba 2 métodos de enseñanza diferentes para aprender anatomía: para determinar si existe significancia educativa se tendría que hacer un estudio cualitativo, encontrando que quizá uno de ellos beneficiará más que el otro; en este caso, no se buscaría tal cual un valor *p*, sino observar en qué grupo de alumnos se obtienen mejores resultados de aprendizaje.

## CONCLUSIONES

A pesar de que la investigación cuantitativa es falsamente considerada la más precisa y válida, su interpretación aún tiene múltiples errores, ya que –como se ha expuesto– la visión reduccionista del valor *p* hace que estudiantes, docentes e investigadores se limiten a ver y “analizar” un simple número con naturaleza probabilística, mas no la posible realidad del fenómeno.

Si bien por el simple hecho de que este trabajo hable de estadística inferencial ya lo convierte (hasta cierto punto) en un artículo positivista, deseamos que la nueva enseñanza en esta área permita formar nuevos profesionales e investigadores con visión holística y, así, terminar con el fomento del reduccionismo e inductivismo ingenuo. Asimismo, consideramos totalmente prudente enfatizar desde la formación a nivel licenciatura en metodologías cualitativas y no solo en las cuantitativas, esto con el objetivo de brindar más armas para el adecuado análisis de las diferentes áreas de estudio dentro de las ciencias médicas.

## CONTRIBUCIÓN INDIVIDUAL

FPS: Idea original, desarrollo teórico, redacción y aprobación del artículo.

## AGRADECIMIENTOS

Ninguno.

## PRESENTACIONES PREVIAS

Ninguna.

## FINANCIAMIENTO

Ninguno.

## CONFLICTOS DE INTERESES

Ninguno. 🔍

## REFERENCIAS

1. Pérez J. El Positivismo y la Investigación Científica. *Rev Empres.* 2015;9(3):29-34. Disponible en: <https://editorial.ucsg.edu.ec/ojs-empresarial/index.php/empresarial-ucsg/article/view/20>
2. Chalmers AF, (aut.), Pérez E, (trad.). ¿Qué es esa cosa llamada ciencia? Una valoración de la naturaleza y el estatuto de la ciencia y sus métodos. 2a ed. Ciudad de México: Siglo XXI Editores; 2019.

3. Kremer-Marietti A. El Positivismo. 2a ed. México: Publicaciones Cruz; 1989.
4. Rodríguez MI, Albert A, Agnelli H. Pruebas de hipótesis y el valor  $p$ : usos e interpretaciones. En: Buendía G, Castañeda A. (eds.). Memoria de la XII Escuela de Invierno en Matemática Educativa. Ciudad Madero: Universidad de Los Andes; 2009. 104-117. Disponible en: <http://funes.uniandes.edu.co/16270/>
5. Daniel WW. Prueba de hipótesis. En: Daniel WW. (aut.), León F. (trad.). Bioestadística. Base para el análisis de las ciencias de la salud. 4a ed. México: Limusa Wiley; 2013. p. 204-94.
6. Romero N. La revolución en la toma de decisiones estadísticas: el  $p$ -valor. Telos. 2012;14(3):439-446. Disponible en: <http://www.redalyc.org/articulo.oa?id=99324907004>
7. Daniel WW. Estadística descriptiva. En: Daniel WW. (aut.), León F. (trad.). Bioestadística. Base para el análisis de las ciencias de la salud. 4a ed. México: Limusa Wiley; 2013. p. 15-56.
8. Daniel WW. Estadística no paramétrica y de libre distribución. En: Daniel WW. (aut.), León F. (trad.). Bioestadística. Base para el análisis de las ciencias de la salud. 4a ed. México: Limusa Wiley; 2013. p. 658-736.
9. Candia R, Caiozzi G. Intervalos de Confianza. Rev Med Chile. 2005;133:1111-1115. doi: 10.4067/S0034-98872005000900017
10. Flores-Ruiz E, Miranda-Navales MG, Villasis-Keever MÁ. El protocolo de investigación VI: cómo elegir la prueba estadística adecuada. Estadística inferencial. Rev Alerg Mex. 2017;64(3):364-370. doi: 10.29262/ram.v64i3.304
11. Tovar CM. El falsacionismo de Popper y sus objeciones al marxismo. Letras. 2019;90(131):210-228. doi: 10.30920/letras.90.131.9
12. Carvalho J, Patino CM. O que realmente significa o valor  $p$ ? J Bras Pneumol. 2015;41(5):485. doi: 10.1590/S1806-37132015000000215
13. Kline RB. Research Trinity. Design, measurement and analysis. En: Kline RB. (aut.). Becoming a behavioral science researcher: a guide to producing research that matters. 2a ed. Nueva York: The Guilford Press; 2020. p. 14-47.
14. Williamson J. Establishing Causal Claims in Medicine. Int Stud Philos Sci. 2019;32(1):33-61. doi: 10.1080/02698595.2019.1630927
15. The RECOVERY Collaborative Group. Dexamethasone in Hospitalized Patients with Covid-19. N Engl J Med. 2021; 384:693-704. doi: 10.1056/NEJMoa2021436