



Cómo entender la investigación: esas temibles estadísticas

La mayoría de los autores que publican artículos de investigación usan estadísticas para sacar sus conclusiones. ¡Agarrarse fuerte! Las estadísticas tienen la forma de perder incluso a los mejores lectores. Lo que trataremos de hacer aquí es brindarle una comprensión muy simple y ágil de las estadísticas.

En general, se usan las estadísticas para *describir* algo o examinar *diferencias* entre grupos o *relaciones* entre características. Los estadísticos usan términos como *media*, *mediana* y *desviación estándar*.

Media:

La media es una palabra formal para denominar al promedio. Es la suma de *todos* los valores o puntajes, dividida por la *cantidad* de personas en el estudio o grupo.

Mediana:

La **mediana**, por otro lado, es el puntaje o valor que está más cerca del *medio*. La mitad de los puntajes individuales son más altos que el mediano mientras que la otra mitad es más baja. Por ejemplo, si tiene cinco números: 0, 0, 5, 10, 30, la **mediana** o promedio sería 9. ($0 + 0 + 5 + 10 + 30 = 45$; $45/5 = 9$). La **mediana**, no obstante, sería 5, porque hay dos puntajes superiores y dos inferiores.

Es bastante obvio que una **media** o promedio podría utilizarse, pero ¿por qué nos interesan las **medianas**? De hecho, los investigadores utilizan la mediana para brindarle al lector más información sobre la media. Aunque la **media** nos dice lo que es el *promedio*, **la mediana** nos dice más sobre lo que es *el típico*. Considere este ejemplo: El precio *promedio* de un hogar en su ciudad puede ser de \$250.000. Esto es en parte porque el menor precio del hogar es de \$90.000 (no hay en realidad hogares que cuesten \$10, \$15, o incluso \$500 para que figuren en el promedio) y en parte porque una pequeña cantidad de hogares multimillonarios impulsan la **media** o el promedio. En un caso como este, la **mediana** puede indicarle que a pesar del alto *promedio*, el "término medio", el típico hogar mediano es más del monto de \$120.000.

El contenido de este documento es sólo para fines de información general y no sustituye el asesoramiento médico profesional o el tratamiento para la afección médica específica. No se insinúa ni establece de otro modo una relación profesional por el hecho de leer este documento. Usted no podrá usar esta información para diagnosticar o tratar un problema de salud o una enfermedad, sin consultar con un proveedor calificado de atención médica. Muchos de los recursos mencionados no están afiliados a Craig Hospital. Craig Hospital no asume responsabilidad por materiales de terceros u otras acciones u omisiones como resultado de contenidos o sugerencias realizadas en este documento, y no se deberá recurrir al mismo sin investigación independiente. La información de esta página es un servicio público suministrado por Craig Hospital y de ninguna manera representa una recomendación o autorización de Craig Hospital.

Desviación estándar:

Otra estadística que verá en la investigación científica es la **desviación estándar**. Le indica el grado de separación entre los datos. Por ejemplo, imagine que va a someterse a una cirugía de columna relativamente nueva. Le dijeron que solo cinco personas tuvieron esa cirugía en su hospital, y que la estadía **media** (promedio) fue de 24 días. ¿Eso es todo lo que necesita saber para decirle a su empleador, o para planificar las necesidades de cuidado personal? De ninguna manera. Piénselo: Suponga que las estadías de hospital para esas cinco personas fueron de 22, 22, 24, 25 y 26 días. La estadía **media** es de 24 días. Suponga que, en cambio, las estadías hayan sido: 6 días, 8 días, 10 días, 31 días, y 65 días. La **media** nuevamente es de 24 días. En el primer ejemplo los valores están muy cerca uno del otro, con un solo día de diferencia. En el segundo ejemplo, los puntajes están más separados, con muchos días entre cada par de puntajes. Esta separación entre puntajes o valores es lo que describe la **desviación estándar**.

Todo lo que describimos hasta ahora recae en el área de las estadísticas *descriptivas*. La idea es darle buena información sobre cómo son los sujetos y datos. ¿El sujeto típico se parece en algo a usted? ¿Las medias y medianas son similares? ¿Los hallazgos están muy separados entre sí?

Mostrar relaciones: *significancia estadística*

El siguiente nivel de estadísticas intenta mostrar las *diferencias* entre dos o más grupos, o *relaciones* entre dos o más cosas diferentes. Suponga, por ejemplo, que un investigador desea demostrar que un nuevo medicamento es efectivo para reducir la cantidad de infecciones de vejiga. Incluso si los que toman el medicamento parecen tener menos infecciones que quienes no lo toman, siempre existe la posibilidad de que *haya* una mejora por casualidad. También siempre existe la posibilidad de que algo se produce por algún factor *que no* es lo que estudiaba el investigador. Para reducir los efectos de estas posibilidades, los estadistas hacen dos cosas:

1. Evaluar sus hallazgos (medias y promedios de los que hablábamos antes) para conocer la **significancia estadística**.
2. Intentar *controlar* otros factores ajenos a los que estudian.

Cuando los investigadores evalúan la importancia estadística, comparan diferentes grupos de valores (infecciones de vejiga antes y después de usar el medicamento) mientras consideran cuántas personas participaron de la investigación, cuán drásticos parecen los hallazgos, y cuáles son algunas de las características de las personas que compararon. Luego usan unas complicadas fórmulas matemáticas para calcular valores de *probabilidad*. Para nuestro medicamento de infección de vejiga, este valor de probabilidad nos dice qué probabilidad hay de que las personas del estudio mejoren porque el medicamento hizo efecto, o si mejoraron simplemente por casualidad o por otro factor desconocido que no fuera el medicamento.

Investigación: estadísticas

Si el investigador descubre que el valor de probabilidad es bajo (generalmente, menos del 5%, 1% o incluso 1/10 de 1%), puede llegar a la conclusión de que el medicamento realmente funciona. Estos valores de probabilidad – llamados valores **p** – representan porcentajes, pero en general están representados como $p < .05$, $p < .01$, o $p < .001$, o más específicamente como $p = .023$, y $p = .0067$. Por ejemplo, $p < .01$ significa que existe menos del 1% de probabilidad de que nuestro medicamento para la vejiga funcione por pura casualidad. Si las probabilidades son bajas, los investigadores las describen como **estadísticamente significativas**. Estas son las palabras clave que debe buscar al leer. Recuerde: Cuanto más *bajo* es el valor **p**, más pequeño es el porcentaje y *mayor* es la importancia y *menos* probable es que algo haya sucedido por casualidad.

¿Cuál es el grado de significancia *suficiente*? Depende de lo que se estudie, el beneficio o daño potencial de los hallazgos, y las propias normas del autor. Si ese medicamento para la vejiga es relativamente económico y seguro y su infección de vejiga no es de una gravedad que atente contra su vida, la pequeña posibilidad de que el medicamento no funcione realmente en forma tan eficaz como indicaron los investigadores (en otras palabras, *significancia* de $p < .05$, o probabilidad de menos de 1 de cada 20 de que funcione) tal vez el tema no sea grave. Si en cambio ese medicamento es para el tratamiento del cáncer y causa severos efectos secundarios, querrá estar seguro de que es extremadamente probable que funcione, antes de tomarlo, o tal vez una significancia de sólo $p < .001$ sea suficiente.

Mostrar relaciones: control

Luego está el tema del *control*. Los investigadores pueden evaluar otros factores que pueden afectar sus resultados. Sigamos analizando nuestro estudio de infección de vejiga. ¿Los investigadores saben antes de empezar si la edad o sexo de la persona tiene efecto en cómo reaccionarán al medicamento? ¿Qué pasa con el tipo de programas que usan los participantes del estudio para controlar sus vejigas? ¿O cuánto líquido beben o si toman vitamina C? Para tratar temas como estos, los investigadores recogen información adicional de cada participante del estudio. Luego incluyen esta información en sus análisis estadísticos y saben si sus hallazgos aún tienen asidero a la luz de estos factores posiblemente complicados. En sus artículos, los investigadores en general le dirán qué factores *evaluaron*. Si *usted* puede encontrar factores que pueden haber tenido un impacto que el investigador no evaluó, usted probablemente interprete sus hallazgos con un poco más de precaución.

“Prueba”

Siempre tenga en mente que la **prueba** es una palabra muy fuerte en la estadística. Los investigadores (especialmente en investigaciones relacionadas con salud y discapacidad) no le dirán en general que han *comprobado* que una cosa causa o lleva a otra. En cambio el mensaje más probable será que una cosa está *relacionada* con la otra. Tenga cuidado. Es fácil engañar sin intención (a veces, intencionalmente) al lector, en especial, cuando solo se da una pequeña cantidad de información. Un ejemplo clarificará esto. Descubrimos que entre sobrevivientes de SCI de edad avanzada, las personas que bebían café en algún momento habían tenido más dolor de hombros tres años después. Esto parece la situación donde una cosa causa la otra; las dos están incluso separadas en el tiempo. Pero esto sigue siendo solamente una *relacion*: beber café de algún modo se *relaciona* con el dolor de hombros. Este hallazgo podría haber aparecido por una cantidad de motivos: tal vez las personas que beben café son más activas, tal vez su

Investigación: estadísticas

sensación de dolor aumenta; tal vez existe alguna otra conexión que no hemos descubierto. O esto puede simplemente ilustrar que a veces hay hallazgos raros. Recuerde: cuando hablamos de cómo $p < .05$ significa que existe un 5% de posibilidad de que el hallazgo se produzca meramente por casualidad. ¡Bien, por lo que sabemos, esta relación entre café y dolor puede caer en ese 5%! Por eso con datos realmente importantes (más como el ejemplo del medicamento para cáncer que usamos anteriormente, que con el tema del café que describimos aquí) los investigadores buenos *repiten* sus hallazgos. Realizan proyectos de investigación completos por segunda vez, con la misma metodología, para ver si obtienen los mismos resultados.

Validez

El último concepto para tratar aquí se llama **validez**. Se explica mejor con un ejemplo.

El investigador John Q, de la universidad nacional más importante, informó que su último estudio demostró una **significativa relación** entre el nivel de lesión de la médula espinal y la inteligencia: Las personas con paraplejía, dijo, son mucho más inteligentes que las personas con cuadriplejía. ¿Cómo llegó a este hallazgo? Usó el test de inteligencia más popular y el mejor disponible. Todos sus sujetos eran de la misma edad y todos tenían el mismo nivel de educación. Todos, por supuesto, hablaban inglés y podían leer fácilmente las preguntas del test. Siguió las reglas del test para administrarlo, y evaluó a todos juntos en la misma aula para asegurarse de que las situaciones de evaluación fueran iguales. Tuvo ayudantes para asegurar que nadie hiciera trampa. Todos tuvieron dos horas para terminar el test. John Q. los corrigió y analizó, y seguramente, esos cuadripléjicos no estaban a la altura de las circunstancias. Su conclusión de que las personas con cuadriplejía no son tan inteligentes llegó a todos los periódicos e hizo que apareciera en algunos programas de televisión.

Pero... existe un pequeño punto que olvidó. ¿Puede adivinar cuál es? El test estaba *cronometrado*. ¿Puede la gente con cuadriplejía escribir tan rápido como los parapléjicos? No. Algunos no pueden escribir en absoluto. Así que, la verdad, es muy probable que no fuera *su cerebro* el responsable de los malos puntajes del test, sino sus *brazos y manos*. Estamos siendo muy petulantes en esto, pero es un concepto muy importante. *John Q. no estaba realmente midiendo lo que creía que medía*. Creyó que estaba evaluando la inteligencia, pero estaba evaluando la velocidad de escritura. Su investigación *no fue válida...* no midió lo que él dijo que medía. ¿Cómo podría haber hecho válida esta investigación? Podría haber dado a todos sus sujetos un test oral; podría haberles dado a todos más tiempo. Tenga en cuenta este concepto, como también todos los demás que describimos antes... y podrá encontrar la forma de criticar los informes de investigación que escuche y lea.

Este folleto apareció originalmente publicado como artículo en la edición de 2001 de *SCI Life* (páginas 32-33) titulado "Como entender la investigación: Parte III: ¡No permita que las estadísticas lo aterren!" Se vuelve a editar aquí con el permiso de la Asociación Nacional de Lesiones de la Médula Espinal. Este folleto se redactó con fondos del Instituto Nacional sobre Discapacidad e Investigación de Rehabilitación del Departamento de Educación de EE. UU. Las opiniones que aparecen en esta publicación son del beneficiario del subsidio y no necesariamente reflejan las opiniones del Departamento de Educación de EE. UU.