

EL PAPEL DEL ANALISIS ESTADISTICO EN LA INVESTIGACION PSICOLOGICA

Enerio Rodríguez Arias

Resumen

Se cuestiona el carácter relativamente autónomo del análisis estadístico en la investigación psicológica, así como el uso rutinario y acrítico que se hace en psicología de los principales procedimientos de análisis estadístico. Se describe la estrategia fisheriana conocida como "prueba de la hipótesis nula", y se ilustra la controversia actual en torno a las pruebas de significación estadística. Se insiste en la importancia de distinguir entre la significación estadística y la significación práctica de los resultados de una investigación. Se ofrece una descripción de la alternativa ofrecida por la estrategia estadística bayesiana, cuyo valor para la investigación psicológica es reconocido, pero cuya escasa difusión en psicología es explicada a la luz de la formación estadística que reciben los psicólogos. Finalmente se insiste en la necesidad de reconsiderar la filosofía subyacente a las prácticas de análisis estadístico vigentes en la investigación psicológica, así como de orientar la enseñanza de la estadística dentro de la psicología más hacia la comprensión del razonamiento estadístico que a la mecánica de los procedimientos de análisis.

Si las relaciones que se dan entre los fenómenos de la naturaleza fueran transparentes, no dependeríamos tanto del análisis estadístico, y quizás no sería tan urgente la invitación a reflexionar y discutir sobre el papel del análisis estadístico en la investigación psicológica. Pero la transparencia de las relaciones es un caso excepcional, por lo que habitualmente, armados con algún instrumento de análisis estadístico, hemos de abrirnos paso a través de cuadros visualmente inescrutables a causa de su complejidad, tras la búsqueda de un sentido en los datos resultantes de nuestras observaciones.

Posiblemente no exista un solo psicólogo que nunca se haya visto envuelto en algún proyecto de investigación, y es casi seguro que todo psicólogo es en mayor o menor grado un lector de investigaciones psicológicas publicas. Y en ese doble papel de productor y/o consumidor de in-

vestigaciones, el psicólogo no debe permanecer ajeno a los problemas que suscita el análisis estadístico. Ante esta realidad, creo que puede ser de utilidad para los psicólogos la exposición de algunas consideraciones sobre el papel del análisis estadístico en la investigación psicológica, lo cual hago como una investigación a la reflexión y discusión sobre el tema.

Es un hecho incuestionable que los principales procedimientos de análisis estadístico no se desarrollaron para satisfacer las necesidades de la investigación psicológica. Dichos procedimientos surgieron vinculados a otras áreas del conocimiento, y los psicólogos los extrapolaron a otras áreas de investigación. Pero salvo algunas excepciones, no se hizo esfuerzo alguno por re-pensar en la lógica de los procedimientos estadísticos ya elaborados y en las posibles limitaciones de su adecuación a la naturaleza de las observaciones psicológicas. El resultado de esta actitud ha sido la difusión y aceptación, con un carácter aparentemente irreversible, de un conjunto de técnicas de análisis estadístico, la mayoría de las cuales descansan en suposiciones que no se cumplen cuando se aplican a datos resultantes de observaciones psicológicas. De esta manera, el análisis estadístico mantiene un carácter relativamente autónomo en el proceso de la investigación psicológica, en lugar de estar subordinado a las características del problema bajo investigación. Es así como algunos psicólogos se han convertido en expertos en numerología, aplicando sofisticados análisis estadísticos a datos muy toscos e imprecisos. Esta grave distorsión del carácter auxiliar y subordinado del análisis estadístico se ha perpetuado entre los psicólogos y estudiantes de psicología debido en gran parte a la forma rutinaria en que se les entrena en el uso de los procedimientos de análisis. Esto no quiere decir que el psicólogo debe reinventar la Estadística; si así fuese, posiblemente ni siquiera todos los estadísticos estarían en condiciones de usar los procedimientos estadísticos. Más bien, lo que quiero decir es que el psicólogo debe dominar los criterios mínimos indispensables para librarse del más grave de los errores en el análisis estadístico, el error de no usar la técnica de análisis apropiada a la naturaleza de los datos, designado por Schlaifer (Bakan, 1974, p 28) como "el error tipo III". El riesgo de incurrir en este tipo de error aumenta en la medida en que el psicólogo concibe las técnicas de análisis como instrumentos carentes de suposiciones, y que por tanto pueden utilizarse indiscriminadamente dondequiera que las observaciones hayan sido traducidas a números, sin preguntarse cuál es el grado de isomorfismo existente entre dichos números y los fenómenos por ellos representados. Entonces, el análisis estadístico, que debe ser un recurso al servicio de la búsqueda de relaciones entre fenómenos observados, se convierte así en una práctica ritual, mantenida más por la fuerza de la costumbre que por su valor para la investigación. Una recomendación útil para enfrentar el riesgo mencionado es atenerse al mandato de no realizar con los números ninguna operación que no pudiese realizarse con las realidades que los números representan; esto quiere decir que primero debemos determi-

nar qué significan los números a los que han sido traducidas nuestras observaciones, para entonces seleccionar una técnica de análisis apropiada a la naturaleza de dichas observaciones. Suponiendo que el psicólogo sabe cómo evitar el mencionado "error tipo III", pasaré a examinar el momento crucial del proceso de análisis estadístico; me refiero a la decisión sobre la hipótesis nula. Para orientación del lector no familiarizado con el tema, presentaré la información básica sobre la hipótesis nula antes de examinar los aspectos controversiales de la decisión sobre la misma. Cómo surgió la llamada hipótesis nula, y qué papel desempeña en la investigación? Fue el estadístico Ronald Fisher quien creó la hipótesis nula como una estrategia en el análisis de datos. Al contrastar su hipótesis sustantiva con los datos reunidos, el investigador puede verse tentado a concluir prematuramente que su hipótesis ha sido confirmada por los datos. A Fisher le pareció necesario buscar una forma de neutralizar esa tentación. Para ello propuso la hipótesis nula. Se trata de una hipótesis que en su sentido más general dice que no hay relación entre los factores bajo estudio, aunque puede asumir varias formas, tales como: que no hay diferencia entre las medias de dos poblaciones definidas, o que no hay diferencia entre las proporciones de sujetos que en dos poblaciones definidas poseen determinada característica, o que no hay diferencia en los rangos ocupados por los sujetos de dos poblaciones definidas, etc. etc. De esta manera, el análisis estadístico no prueba directamente la hipótesis sustantiva del investigador, sino la hipótesis nula; es decir, sólo después que haya rechazado la hipótesis nula, puede el investigador considerar la posibilidad de que los datos confirmen su hipótesis sustantiva. En el caso contrario, es decir, en el caso de que el análisis estadístico no le permita rechazar la hipótesis nula, debe abstenerse a toda consideración sobre su hipótesis sustantiva. De manera que en la formulación original de Fisher, la única decisión que podía tomar el investigador era rechazar la hipótesis nula, lo cual debía hacer cuando la probabilidad de explicar con dicha hipótesis la relación entre los factores fuese igual o menor que 0.05. Por tanto, en la idea de Fisher, el único error que podría cometer el investigador consistía en rechazar una hipótesis nula verdadera, esto es, descartar el azar como explicación, cuando en realidad los resultados se pueden explicar por azar. Posteriormente, se introdujo la posibilidad de un segundo tipo de error, el cual consistiría en no rechazar una hipótesis nula falsa, esto es, en creer que el azar puede explicar resultados que en realidad son inexplicables por azar. Fue entonces cuando se acuñó el nombre de error tipo I para designar el error mencionado por Fisher, llamándole "error tipo II" a este nuevo tipo de error; y fueron estas designaciones las que indujeron a Schlaifer a designar como "error tipo III" al error previamente mencionado de no utilizar la técnica de análisis apropiada.

No hay duda de que la prueba de la hipótesis nula, tal como fue concebida por Fisher, perseguía darle al investigador cierto grado de garan-

tía para distinguir la realidad de la ficción. La repetición de los resultados de una investigación por parte de investigadores independientes es la mejor garantía de la realidad de dichos resultados. Sin embargo, por razones de una supuesta originalidad, casi ningún investigador reproduce la investigación realizada por otro. De manera que al final, el investigador se ve en la necesidad de decidir si efectivamente él ha encontrado alguna relación entre los factores bajo investigación. Entonces, procede más o menos de la siguiente manera: Supone que los resultados de su investigación constituyen una muestra extraída de una población de resultados posibles, es decir, la población hipotética integrada por los resultados arrojados por todas las investigaciones iguales que podrían hacer otros investigadores. La estrategia de la prueba de la hipótesis nula consiste en comparar el resultado obtenido por el investigador con el resultado que podría ocurrir por azar con una probabilidad de 0.50 en el supuesto de que la hipótesis nula fuese verdadera. Si el resultado obtenido en la investigación tiene una probabilidad de ocurrir por azar igual o menor que 0.05, el investigador decide que la hipótesis nula es falsa. El razonamiento de Fisher es que no hay razón para que nos ocurra un fenómeno teóricamente raro bajo la hipótesis nula. Si el mismo ocurre, podemos concluir que la hipótesis nula es falsa. No voy a discutir ahora si debemos aceptar que un fenómeno que ocurra el 5 o/o de las veces o menos es realmente un fenómeno raro. ¿Por qué el 5 o/o? Cowles y Davis (1982, p. 553) en estudio sobre los orígenes de 0.05 como nivel de significación, reproducen las opiniones de Fisher para justificar esa práctica. En efecto, en la página 47 de la edición en inglés de *Métodos Estadísticos para Investigadores* (1925) aparece la primera mención específica al respecto. Dice Fisher: "Es conveniente tomar este punto (0.05, ERA) como un límite al juzgar si una desviación va a ser considerada significativa o no. De esta manera, las desviaciones que excedan dos veces la desviación standard son formalmente consideradas como significativas". Y poco tiempo después (1926), en un artículo dedicado al diseño de experimentos de campo, Fisher reitera su preferencia por el nivel de 0.05, aunque reconoce que se pueden usar otros niveles. Decía Fisher entonces: "Si uno en veinte no parece una probabilidad suficientemente alta, podemos, si lo preferimos, trazar la línea en uno en cincuenta (el punto del 2^o), o uno en cien (el punto del 1 o/o). Personalmente, prefiero (decía Fisher) establecer un nivel bajo de significación en el punto del 5 o/o, e ignorar totalmente todos los resultados que no alcancen este nivel. Un hecho científico debe considerarse experimentalmente establecido sólo si un experimento adecuadamente diseñado *raramente falla* en alcanzar este nivel de significación". Esta es la práctica vigente, a pesar de recientes cuestionamientos a la misma. Por ejemplo, en su libro *Principios Estadísticos en el Diseño Experimental* (1926, p.13), Winer reproduce parcialmente un comentario del Editor de la Revista *Biometrika* que entre otras cosas dice: "El uso frecuente de los niveles de significación 0.05 y 0.01 es un asunto convencional que

tiene poco fundamento lógico o científico". Igualmente, en una crítica al carácter sagrado del 0.05, Skipper, Guenther y Nass (1967) señalan que "la adhesión ciega al nivel de 0.05 impide la consideración de estrategias alternativas, y es un grave obstáculo para la interpretación de los datos". Además, después de señalar que la dicotomía entre resultados "significativos" y "no significativos" puede confundir por igual a profesionales y legos, sugieren que un enfoque más racional sería reportar el nivel real de significación, y dejar la responsabilidad de la interpretación al lector. Pero estas críticas y sugerencias han tenido poco impacto en las prácticas de publicación de investigaciones científicas.

Después de este largo paréntesis sobre el nivel de significación estadística, volvamos a un aspecto esencial en la estrategia fisheriana de la prueba de la hipótesis nula. Se trata de que en esta estrategia cada decisión sobre la hipótesis nula es tomada como si fuese única; es decir, cada investigación es interpretada como si fuese la primera vez que se realiza, de manera que en este modelo de inferencia el investigador debe proceder como si no conociera los resultados arrojados por investigaciones anteriores realizadas de la misma manera que la suya. La razón de esto es que la validez del modelo de inferencia resultaría viciada, si la decisión sobre la hipótesis nula es influida por la información del investigador sobre resultados anteriores. Resulta evidente que esta estrategia desalienta la repetición de una investigación. En una reducción al absurdo de esta posición, Bakan (1974, p.5) ilustra cómo podría resultar confirmada la hipótesis sustantiva de que cada moneda posee un espíritu, el cual, si se le implora adecuadamente, puede hacer que la moneda caiga cara o escudo. Bakan dice que él le pidió al espíritu que hiciera que la moneda cayera cara. La arrojó sucesivamente en seis ocasiones, y en cada ocasión la moneda cayó cara. Bajo la hipótesis nula, la probabilidad de que una moneda caiga cara seis veces seguidas es $(0.5)^6 = 0.015$; por consiguiente, este resultado es significativo al nivel de 0.05. Se rechaza la hipótesis nula, y como la moneda no estaba arreglada para caer cara ni hubo fraude al momento de tirarla, se concluye que la hipótesis de que la moneda posee un espíritu ha sido confirmada. No se puede objetar que nadie ha encontrado ese mismo resultado ni antes ni después de Bakan, pues la lógica del modelo de inferencia de la prueba de la hipótesis nula no lo exige. Aunque Bakan ofrece el ejemplo para ilustrar el absurdo a que puede conducir este modelo de inferencia, me parece que el ejemplo ilustra también cómo el simple análisis estadístico no puede confirmar ninguna hipótesis sustantiva. La conclusión de que los resultados de una investigación sustentan determinada hipótesis sustantiva depende de un razonamiento que no puede ser sustituido por ningún análisis estadístico. Para no ser injusto con Fisher, debo distinguir entre la creencia personal de Fisher y el modelo de inferencia a que se suscribió. Pues bien, Fisher (citado por Edwards, 1972, p.21) decía que "para afirmar que un fenómeno natural es experimentalmente demostrable necesitamos, no un registro aislado,

sino un método confiable de procedimiento. En relación con la prueba de significación, podemos decir que un fenómeno es experimentalmente demostrable cuando sabemos cómo realizar un experimento que raramente falle en darnos un resultado estadísticamente significativo". Sin embargo, esta creencia no aparece integrada al modelo de inferencia elaborado por Fisher.

En vista de la gran importancia que se da a la prueba de la hipótesis nula en las investigaciones psicológicas, conviene pasar revista rápidamente a sus principales deficiencias, y a las formas en que las mismas repercuten en el proceso de investigación. Lo primero que hay que señalar es que esta estrategia, a pesar del propósito original de Fisher, está prejuiciada en contra de la hipótesis nula, pues no hay ninguna razón válida para esperar que ésta sea verdadera en alguna población (Edwards, Lindman, & Savage, 1963). Una prueba de esto es que con muestras muy grandes, todas las hipótesis nulas resultan rechazadas. Este hecho ha llevado a Nunnally (1960, p. 643) a decir que "si la hipótesis nula no es rechazada, es porque N es demasiado pequeña. Si se reúnen datos suficientes, la hipótesis nula será generalmente rechazada. Si el rechazo de la hipótesis nula fuese la intención real en las investigaciones psicológicas, entonces no habría necesidad de reunir datos". Por ejemplo, con una N de 700, hasta una correlación de 0.08 es significativa más allá del nivel de 0.05, y correlaciones sin ninguna importancia, ni teórica y práctica, resultan estadísticamente significativas si la muestra es muy grande. Por otra parte, el prejuicio de esta estrategia en contra de la hipótesis nula es tal que ésta resulta rechazada aun cuando muchos resultados hablan a su favor más que en su contra. Tal cosa ocurre con los resultados que son estadísticamente significativos en la frontera misma de la significación. En estos casos, la hipótesis nula es rechazada porque p es ligeramente menor que 0.05. Ahora bien, si la hipótesis nula fuese realmente falsa, con el aumento en el número de pruebas realizadas debería haber una proporción sustancial de rechazos más dramáticos que de rechazos fronterizos (Bakan, 1974, p.6). Por otra parte, como el resultado que permite rechazar la hipótesis nula es considerado más importante que el que no lo permite, y como los editores de revistas no pueden publicar todas las investigaciones que les llegan, las revistas psicológicas se ven prácticamente inundadas de investigaciones en que la hipótesis nula es rechazada, y casi nunca aparece el caso contrario. Veamos con un ejemplo hipotético hasta dónde podría conducirnos esta práctica. Supongamos que un hipótesis nula es verdadera y que cien investigadores la someten a prueba; 95 de ellos no pueden rechazarla, pero hay 5 que sí logran hacerlo. De estos 95, casi ninguno se atrevería a someter sus resultados para publicación, y si alguno lo hace, es probable que no consiga su propósito. En cambio, los cinco que lograron rechazar la hipótesis nula (en este caso, cometiendo el error tipo I, porque se trata de una hipótesis nula verdadera) no encontrarían ninguna dificultad para ver publicados sus resultados. Lo que quiero indicar con este ejemplo es que el prejuicio en contra de la

hipótesis nula ha llegado tan lejos que es muy probable que en las investigaciones publicadas haya una proporción mucho mayor de error tipo I que el 5 o/o que generalmente se supone como riesgo.

Una segunda deficiencia de la prueba de la hipótesis nula como modelo de inferencia es que induce al investigador a despreciar la magnitud real del efecto producido por un factor. Sólo así se comprende que un grupo de 19 investigadores en psicología consultado por Rosenthal y Gaito (1963) respondiesen que tenían más confianza en los resultados que rechazan una hipótesis nula cuando dichos resultados eran obtenidos en una muestra de 100 casos que cuando eran obtenidos en una muestra de 10 casos. Y la realidad es que para rechazar una hipótesis nula con una muestra de 10 casos se necesita un efecto más dramático que el necesario para rechazarla con una muestra de 100 casos. Ahora bien, si lo que importa es rechazar la hipótesis nula, sin consideración a la importancia teórica o práctica de la investigación, entonces no hay duda de que para ese propósito merece más confianza una muestra de 100 casos que una de 10.

La tercera deficiencia de la prueba de la hipótesis nula es que trata la aceptación o el rechazo de una hipótesis como si fuese un proceso automático de toma de decisión. El propio Fisher (1955, p. 74) creía haber logrado una especificación relativamente completa del proceso de inducción, lo cual permitiría lograr la automatización de la inferencia inductiva. En este sentido, decía "Durante siglos se ha sabido que el proceso inductivo existe y es posible para las mentes normales; pero es sólo con el desarrollo reciente de la ciencia estadística que se ha podido hacer una descripción analítica del mismo, casi tan satisfactoria y completa como la que se ha hecho tradicionalmente de los procesos deductivos". La consecuencia de este intento de automatizar el proceso de la inferencia inductiva es que el investigador se ve forzado a rechazar o no rechazar la hipótesis nula si el resultado obtenido cae dentro o fuera de una región de rechazo previamente establecida. No hay grados en esta decisión; se trata de todo o nada; de una separación entre la realidad y la ficción. Lo curioso es que el criterio que regula esta decisión tan rígida ha sido establecido de un modo relativamente arbitrario; me refiero a la probabilidad adoptada como nivel de significación. Una hipótesis sólo puede ser totalmente rechazada o aceptada en el caso improbable de que sea lógicamente incompatible con los datos o lógicamente deducible de ellos. En la realidad, el proceso inferencial no es un asunto de aceptar o de rechazar una hipótesis, es más bien un ajuste en el grado en que uno acepta o cree en una hipótesis a la luz de los resultados arrojados por una investigación. La aceptación o el rechazo de una hipótesis no es una decisión que se toma, es más bien un proceso cognoscitivo, un aumento o una disminución en el grado en que uno cree en dicha hipótesis. La prueba de la hipótesis nula limita arbitrariamente la significación del resultado de cualquier investigación a las opciones de una disyunción excluyente, rechazo o no rechazo de la hipótesis nula. Además establece

una transición discontinua entre las opciones, utilizando un punto de transición arbitrario. Según Rozeboom (1960, p. 428) "el principal error de la prueba de la hipótesis nula reside en creer que la meta de una investigación científica es una decisión y no una evaluación cognoscitiva de proporciones". Recientemente, Smith (1983, p. 317) puso de manifiesto el absurdo que puede derivarse de la concepción de la prueba de la hipótesis nula como proceso automático de toma de decisión. Resulta que en ocasiones un resultado que no es estadísticamente significativo está mucho más cerca del primer valor significativo que del valor 0.00 teóricamente predicho por la hipótesis nula. Sin embargo, la rigidez del procedimiento obliga al investigador a decidir como si su resultado fuese igual a 0.00. En este sentido, dice Smith "cuando sobrevaloramos la prueba de significación hasta el punto que desalentamos la publicación de resultados casi significativos, estamos actuando como si hubiese algo mágico en nuestra arbitrariedad. Otro enfoque consistiría en que las revistas publiquen investigaciones bien hechas, al margen de si alcanzan "significación" o no; esto les daría a los lectores informados la oportunidad de evaluar por sí mismos la importancia (de una investigación)".

Frente al culto a la significación estadística y sus graves consecuencias para la investigación psicológica, Edwards (1972) sugirió la idea de "significación práctica de un resultado, para la cual se toma en cuenta el significado, en algún sentido práctico, de la magnitud en que dicho resultado se desvía de la hipótesis nula, junto con el número de observaciones realizadas al obtener la significación estadística. Un investigador en el campo de la psicología clínica (Garfield, 1978) deplora el énfasis exclusivo que ponen los investigadores de esta área en la significación estadística y su desprecio por la significación psicológica práctica de sus resultados, y añade que estos parecen haber sido conducidos a la adoración en el templo de la significación estadística. Para determinar la significación práctica de un resultado es importante saber qué proporción de la varianza es explicada en función de los factores o variables bajo investigación. Por ejemplo, elevar al cuadrado un coeficiente de correlación nos dice la proporción de la varianza compartida por los factores correlacionados. De igual manera, el cuadrado de la razón de correlación (eta cuadrada) nos dice qué proporción de la varianza total es explicada por los factores o variables independientes estudiadas. Así puede determinarse qué correlaciones bajas que poseen alta significación estadística, carecen totalmente de significación práctica, pues apenas explican el 1 o/o de la varianza. Algo similar ocurre con algunas F que son estadísticamente significativas al nivel de 0.05, y sin embargo la eta cuadrada revela que sólo explican menos del 20 o/o de la varianza, por lo que el resultado, aunque estadísticamente significativo, posee escasa significación práctica, pues deja sin explicar más del 80 o/o de la varianza. Algo similar ocurre con los instrumentos de diagnóstico psicológico. Aunque para los fines de la significación estadística, podría bastar con saber que una técnica de diagnóstico permite distinguir dos o más grupos clínicos con un nivel de significación de 0.05,

para la significación práctica de dicha técnica debemos conocer no sólo la cantidad de casos diagnosticados o clasificados correctamente, sino también la cantidad de diagnósticos o clasificaciones equivocadas, es decir, los llamados falsos positivos y negativos; un número muy elevado de éstos últimos revelaría el escaso valor práctico de la técnica en cuestión.

Frente a las dificultades de la estrategia inferencial fisheriana previamente documentadas, algunos psicólogos han elaborado una estrategia inferencial que toma como su punto de partida el teorema Bayes. Dicho teorema fue publicado originalmente por Thomas Bayes en 1763 bajo el título de "*Ensayo hacia la solución de un problema en la doctrina del azar*", y fue reimpresso en 1958 por la revista *Biometrika*. La primera propuesta para desarrollar una inferencia estadística bayesiana en la investigación psicológica fue hecha por Edwards, Lindman y Savage en 1963. La receptividad de los psicólogos a la estadísticas bayesiana ha sido muy lenta, a pesar de que para los problemas de investigación psicológica parece ser mucho más apropiada que la estadística fisheriana. Aunque existen varios textos generales de Estadística escritos desde la perspectiva bayesiana, y uno que otro aplicado a campos específicos, no existe ninguno aplicado a la Psicología, aunque un libro reciente dedicado a problemas estadísticos y metodológicos de la investigación en Psicología y Ciencias Sociales dedica tres de sus doce capítulos a la Estadística Bayesiana (Karen 1982).

Reduciéndolo a sus dimensiones más simples, podemos decir que el teorema de Bayes no es más que un procedimiento para encontrar probabilidades condicionales, y puede derivarse fácilmente de la definición usual de probabilidad condicional. La forma más simple del teorema de Bayes implica la probabilidad condicional de un fenómeno individual (Winkler, 1982). Creo que un ejemplo facilitaría mucho la comprensión de la idea central del teorema de Bayes. Un psicólogo clínico podría razonar de la siguiente manera: La probabilidad de que este paciente se suicide depende directamente de la probabilidad de que amenace con hacerlo en una circunstancia específica (por ejemplo, deprimido porque fue despedido del trabajo), e inversamente de la probabilidad de que amenace con hacerlo en cualquier circunstancia, es decir, al margen de una circunstancia específica. En este caso, la probabilidad de suicidio es una razón entre dos probabilidades y consiste en pesar una probabilidad contra otra. El teorema de Bayes permite determinar cómo la reunión de información adicional sobre el paciente podría modificar la probabilidad antecedente (Rychlak, 1981, p. 154). Veamos cómo funciona el razonamiento bayesiano en la prueba de hipótesis. Para empezar, el estadístico bayesiano no se pregunta si la hipótesis es verdadera o falsa, sino cuál es la probabilidad de que sea verdadera a la luz de las evidencias disponibles, y cómo se interpreta esa probabilidad. Su respuesta es que la probabilidad puede variar de 0 a 1, y que se interpreta subjetivamente en tér-

minos de probabilidad personal. Mientras la estrategia fisheriana es asimétrica, en el sentido de que los datos sólo hablan para rechazar la hipótesis nula, en la estrategia bayesiana los datos pueden tanto debilitar como fortalecer la hipótesis nula. Al comparar la prueba de hipótesis por los métodos fisheriano y bayesiano, Pitz (1982, pp. 278-279) señala que "la razón de probabilidad arguye contra la hipótesis nula con menos fuerza que el nivel de significación. De hecho, cuando se trabaja con muestras grandes, los datos que conducen al rechazo de la hipótesis nula en niveles pequeños de significación frecuentemente apoyan a la hipótesis nula de acuerdo con la razón de probabilidad. Dado que los procedimientos tradicional y bayesiano conducen en estos casos a conclusiones totalmente opuestas, por lo menos uno de ellos debe estar equivocado. Si uno está interesado en la probabilidad de que la hipótesis nula sea verdadera después que se han reunido los datos, la única respuesta correcta la proporciona la solución bayesiana".

En un argumento a favor del esquema bayesiano de inferencia, Salmon (citado por Weimer, (1979, p. 138) señala que hay resultados positivos que no confirman una hipótesis, es decir, que no añaden nada a su grado de confirmación basado en evidencias anteriores. Esto significa que no toda evidencia que es compatible con una hipótesis constituye un apoyo empírico para la misma. El modelo bayesiano tiene una forma de reconocer esto, pues habla de probabilidades antecedentes de las hipótesis, y en tales casos, la nueva evidencia no transformaría la probabilidad antecedente de una hipótesis en una probabilidad posterior más alta.

No hay duda de que el modelo bayesiano refleja con más fidelidad las creencias y conductas de los investigadores. Prácticamente ningún investigador realiza una investigación sin alguna expectativa concreta basada en conocimientos antecedentes, y prácticamente ningún investigador hace depender totalmente la aceptación o el rechazo de una hipótesis conceptual del resultado de un solo estudio. En términos bayesianos, el investigador comienza cada investigación con algún grado de confianza (un grado de probabilidad subjetiva) en una hipótesis particular, y luego revisa su grado de confianza en esa hipótesis dependiendo de los resultados de la investigación. No hay duda de que esta es una descripción más realista de la conducta de investigación que la ofrecida por el modelo fisheriano de la prueba de la hipótesis nula.

Cómo se explica entonces el desconocimiento general de la estadística bayesiana que reina entre los psicólogos. La respuesta es multidimensional. En primer lugar, hay que reconocer que la estadística fisheriana es el paradigma de la ciencia estadística. Esa es la estadística que se enseña normalmente a todas las profesiones que de algún modo requieren de conocimientos estadísticos mínimos. Esto ha permitido que dicha estadística haya podido ser codificada de manera que su difusión resulte

relativamente fácil. Es posible que nunca la estadística bayesiana alcance la simplicidad de codificación que hoy tiene la estadística fisheriana. En segundo lugar, el costo en esfuerzo humano colectivo que representa el cambio de un procedimiento ya establecido es muy alto, pues no se trata de un simple cambio de fórmulas, sino un cambio en los fundamentos lógicos del análisis. La introducción en el modelo bayesiano de las creencias subjetivas del investigador constituye un elemento incómodo para investigadores acostumbrados a la automatización de los procedimientos. Lo mismo puede decirse de la necesidad de pensar las posibilidades de análisis en función de la naturaleza del problema de investigación, mientras que en la estadística fisheriana un análisis vale como la misma respuesta ante muchas preguntas diferentes, y llevando el razonamiento hasta el absurdo, como la respuesta para una pregunta que nadie ha hecho. Por todo esto, creo que en el presente las posibilidades de una alternativa a la estrategia inferencial fisheriana son remotas, pero esta situación no significa que haya que abstenerse de dar a conocer las principales diferencias de una estrategia de análisis que en la actualidad sobrevive a su valor. Creo que el punto de partida de una renovación de esta fase del proceso de investigación en psicología implica un cambio en la filosofía subyacente al papel del análisis estadístico en el mismo. El investigador no debe delegar en el análisis la responsabilidad de afirmar cuándo las observaciones realizadas hacen aumentar, disminuir o mantener constante su creencia en una proposición.

Los procedimientos de análisis estadísticos sirven en la medida en que ayudan al pensamiento, no cuando se pretende sustituirlo. Berkson (citado por Edwards, Lindman y Savage, 1963, p. 217) ha sugerido el uso de "la prueba traumática interocular, usted sabe lo que significan los datos cuando la conclusión lo golpea entre los ojos". Con esto quiere decir que en ocasiones la inspección directa de los datos puede ser una vía insustituible para interpretar adecuadamente el significado de una investigación.

Los responsables de la formación estadística de los psicólogos podrían contribuir más que nadie a desarrollar una nueva actitud hacia el valor del análisis estadístico en la investigación psicológica. El énfasis excesivo puesto en la rutina de los cálculos, además de limitar el espacio a la discusión de los razonamientos subyacentes, contribuye a desarrollar una actitud de servidumbre frente a la Estadística, y una creencia mágica en la supuesta omnipotencia del análisis estadístico. Sólo así se entiende la actitud del estudiante que renuncia a lo que de manera patente le dicen los datos, para seguir lo que resulta de un análisis aunque le parezca absurdo. Su servidumbre ante la Estadística le lleva a dudar de su propia capacidad de interpretación, pero no del análisis. Esta situación debe ser cambiada, o de lo contrario el análisis estadístico terminará por perder valor para las investigaciones psicológicas. Terminaré esta exposición reiterando el contenido esencial de su mensaje: Es necesario restituir al aná-

lisis estadístico su carácter auxiliar y subordinado en la investigación psicológica. El psicólogo nunca debe olvidar que sus hipótesis, en tanto que hipótesis científicas, no son de la misma naturaleza que las hipótesis estadísticas (Bolles, 1962) que se aceptan o se rechazan por medio de decisiones. Las hipótesis científicas son más bien creencias que se fortalecen o se debilitan gradualmente en función de las evidencias disponibles. Ningún análisis estadístico podrá sustituir al investigador en la tarea de tratar de distinguir la realidad de la ficción, cuya frontera parece ser un continuo oscuro imposible de iluminar por medio de un criterio discontinuo, establecido arbitrariamente.