



Revista Mexicana de Análisis de la Conducta

ISSN: 0185-4534

editora@rmac-mx.org

Sociedad Mexicana de Análisis de la Conducta
México

ACUÑA, LAURA

EL USO DE ESTADÍSTICA EN ANÁLISIS DE LA CONDUCTA: ¿CUÁNDΟ USARLA Y CUÁNDΟ NO?

Revista Mexicana de Análisis de la Conducta, vol. 36, núm. 1, junio, 2010, pp. 131-143

Sociedad Mexicana de Análisis de la Conducta

Guadalajara, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=59313134009>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

EL USO DE ESTADÍSTICA EN ANÁLISIS DE LA CONDUCTA: ¿CUÁNDO USARLA Y CUÁNDO NO?

**THE USE OF STATISTICS IN BEHAVIOR ANALYSIS: WHEN TO USE IT
AND WHEN NOT?**

LAURA ACUÑA*

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Resumen

El análisis de la conducta proviene de una tradición basada en las ciencias biológicas, que favorece las demostraciones de la replicabilidad de un hallazgo analizando directamente la conducta individual. En otras áreas de la psicología basadas en las ciencias sociales se favorece la demostración de la replicabilidad de un hallazgo analizando mediante estadística inferencial la conducta agregada, ya sea de los mismos individuos o de una colectividad de individuos. Aunque los analistas conductuales favorecen el análisis de datos individuales, el número de artículos publicados en revistas especializadas que incluyen estadística inferencial ha aumentado durante los últimos años. En este trabajo se presentan las premisas básicas de ambas estrategias, la basada en las ciencias biológicas y la basada en las ciencias sociales, para establecer la confiabilidad de la replicabilidad de un hallazgo. También se presentan las principales críticas y defensas hechas por los partidarios de cada enfoque. Se concluye que cada estrategia tiene sus propias fortalezas y debilidades y que para determinar la confiabilidad de un hallazgo un investigador debe analizar sus datos usando la estrategia que le permita avanzar el conocimiento al comparar sus hallazgos con los de investigación anterior y responder sin ambigüedad su pregunta de investigación.

Palabras clave: *análisis experimental de la conducta, inspección visual de los datos, estadística descriptiva, estadística inferencial.*

*Enviar correspondencia a la autora: Laboratorio de Condicionamiento Operante, Facultad de Psicología, UNAM, Ave. Universidad 3004, Col. Copilco-Universidad, México, D. F., 04510. Dirección de correo electrónico: lacuna@servidor.unam.mx. El texto de este artículo está basado en la conferencia dictada por el autor durante el XIX Congreso Mexicano de Análisis de la Conducta, Aguascalientes, Ags., marzo 23-25, 2009.

Abstract

Behavior analysis originated from the tradition of the biological sciences, that favor the replication of data analyzing directly the behavior of individual subjects. Other areas of psychology that originated from the social sciences, favor the replication of data analyzing statistically grouped data, either of the behavior of the same subjects or of groups of subjects. Although behavior analysts favor the analysis of individual data, there has been an increase in the number of articles published in specialized journals in which inferential statistics has been used to determine the reliability of a finding. In this paper the basic premises of both strategies to infer the reliability of a finding, that derived from the biological sciences and that derived from the social sciences, are presented. The paper also includes the main criticisms and the arguments against those criticisms made by the defenders of each strategy. The conclusion is that each strategy has its own strengths and weaknesses and that in order to determine the reliability of a finding, a researcher should analyze the data following the strategy that allows him/her to advance scientific knowledge by comparing the current findings with those from previous research and thus answer unambiguously the research question.

Key Words: *experimental analysis of behavior, visual analysis of the data, descriptive statistics, inferential statistics.*

La lógica de la investigación científica exige que la relación entre una variable independiente y una variable dependiente sea replicable. Sin embargo, las ciencias biológicas y las ciencias sociales han abordado de diferente manera la demostración de la replicabilidad del descubrimiento científico. Mientras que disciplinas biológicas como la etología, la fisiología y la psicología han favorecido la replicabilidad del dato en sujetos individuales, las disciplinas sociales en general han favorecido la replicabilidad del dato observando agregados de conductas individuales. Dado que el análisis de la conducta proviene de las ciencias biológicas, es natural que favorezca las demostraciones de replicabilidad observando directamente conductas individuales (Bernard, 1957; Sidman, 1960). De la misma forma, aquellas áreas de la psicología que se relacionan con las ciencias sociales, naturalmente favorecen datos agregados para demostrar la replicabilidad de sus hallazgos (Brown, 1968).

Aunque los analistas conductuales prefieren el análisis de los datos de sujetos individuales, es interesante notar que las revistas especializadas como la *Revista Mexicana de Análisis de la Conducta* (RMAC) y el *Journal of the Experimental Analysis of Behavior* (JEAB) han publicado un número creciente de trabajos que incluyen el análisis de datos agregados mediante estadística.

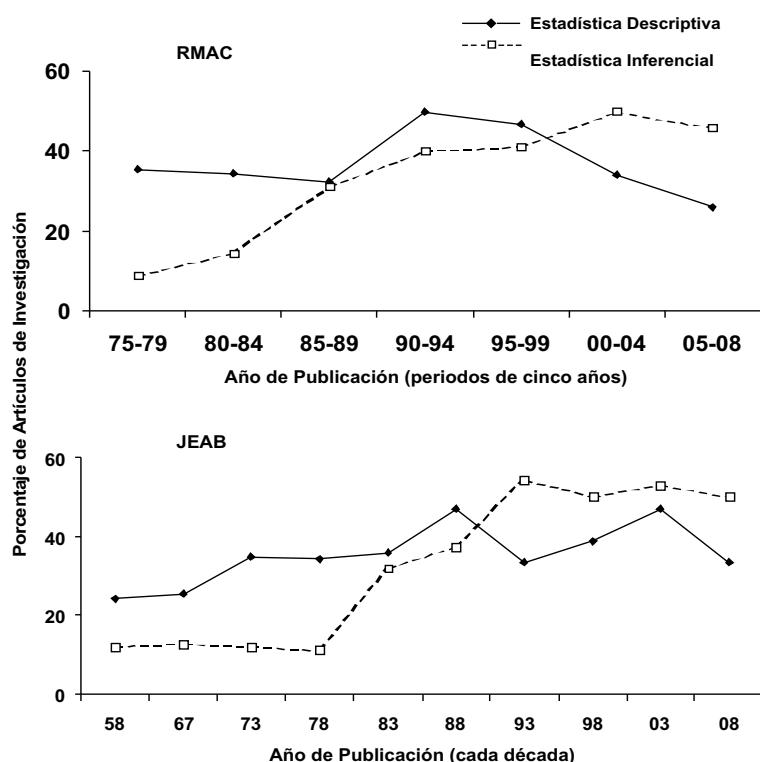


Figura 1. Porcentaje de artículos de investigación publicados en la RMAC y en el JEAB que han incluido análisis estadísticos de los datos.

En el panel superior de la Figura 1 se muestra el porcentaje de artículos de investigación publicados en la RMAC que han incluido análisis estadísticos de datos agregados, utilizando ya sea estadística descriptiva o inferencial. Los porcentajes se muestran agrupados durante cinco años desde el volumen 1 (1975), hasta el volumen 34 (2008). En el panel inferior de la figura se muestra el porcentaje de artículos publicados en el JEAB que han incluido análisis estadísticos descriptivos e inferenciales, de datos agregados. Los datos se muestran para los volúmenes correspondientes a cada década (i.e., volúmenes 1, 10, 20 y así sucesivamente), desde 1958 (volumen 1) y hasta 2008 (volumen 90). Es pertinente enfatizar que hasta la década de los 90's la mayor parte de los trabajos habían utilizado estadística descriptiva y muy pocos estadística inferencial. Este último punto se debe a que los supuestos

en los que se basa la estadística inferencial son controversiales (e.g., Killeen, 2005, 2006; Miller, 2009; Sanabria & Killeen, 2007).

El propósito de este trabajo es reseñar de forma muy general la lógica que subyace al análisis de datos individuales y posteriormente la que subyace al análisis de datos agregados. Para cada método se discuten las principales críticas y argumentos en defensa de dichas críticas. Se concluye con una reflexión sobre cuándo utilizar análisis de datos individuales o de datos agregados. Dado que el trabajo no pretende una revisión exhaustiva de la lógica de cada tipo de análisis ni de todas las críticas formuladas, está dirigido a investigadores no familiarizados con los argumentos básicos del debate sobre el uso de uno de los tipos de análisis de datos.

Análisis de Datos Individuales

La conducta de los organismos individuales es la materia prima de la psicología. La forma más directa para establecer la replicabilidad de alguna variable independiente es examinar el cambio en la variable dependiente directamente, ya sea de una manera fina o en medidas agregadas (Plutchik, 1968). Bajo esta lógica, los investigadores adeptos a las ciencias biológicas favorecen diseños que alternan la intromisión y el retiro de alguna variable independiente (e.g., diseños de línea base, tratamiento y re establecimiento de la línea base) (Skinner, 1938). Otra forma de abordar el problema de la replicabilidad del dato es utilizar la condición previa para evaluar el efecto de la condición subsiguiente (e.g., una serie ascendente o descendente de diferentes valores de la misma variable independiente) (cf. Sidman, 1960).

Críticas al Análisis de Datos Individuales

Por lo general la observación de medidas repetidas obtenidas de sujetos individuales involucran a pocos sujetos (estrategia también conocida como “diseños de un solo sujeto”) (Sidman, 1960). Aquellos investigadores que favorecen el análisis estadístico de datos agregados han criticado el uso de pocos sujetos alegando que un número reducido no garantiza la generalidad de los hallazgos a otros sujetos de la misma población (e.g., Francis, 2005; Newcomb, 2005).

Otra crítica a la observación de medidas repetidas de los mismos sujetos es que los investigadores que se adhieren a esta estrategia frecuentemente se limitan a caracterizar sus hallazgos empleando datos directos (e.g., llamada “inspección visual” de figuras y tablas), ya sea de una manera fina o de forma agregada (e.g., Fish, 2001). Esta forma de caracterizar los datos permite la inferencia de si la variable independiente tuvo o no un efecto confiable (Perone, 1999; Sidman, 1960). No obstante, en algunos estudios se ha documentado que la inferencia sobre el efecto de la variable independiente puede depender

de la persona que mira los datos y por tanto está sujeta a influencias idiosincráticas de cada investigador (e.g., Danov & Symons, 2008; DeProspero & Cohen, 1979). En contraste, los investigadores que favorecen la estadística alegan que deben practicarse análisis imparciales para determinar si tales efectos se deben o no a razones fortuitas (Kazdin, 1982).

Defensa del Análisis de Datos Individuales

Los investigadores que prefieren las medidas repetidas de un mismo sujeto responderían a la crítica concerniente a la generalidad de los hallazgos alegando que el problema de la generalidad de un dato es un problema empírico que no puede solucionarse aumentando el número de sujetos (Sidman, 1960). Específicamente, dirían que para establecer la generalidad de un hallazgo habría que replicar el experimento sistemáticamente variando las características de los sujetos de la misma población (e.g., Branch, 1999; Perone, 1999; Reese, 1998). De hecho, los resultados de investigaciones que incluyen medidas repetidas de la conducta de un mismo sujeto y replicaciones directas y sistemáticas del efecto con otros sujetos de la misma y de diferentes especies han mostrado evidencia de relaciones entre variables independientes y dependientes ordenadas, confiables y de gran generalidad (Shull, 1999).

Con respecto a la discusión sobre la suficiencia del examen de los datos directos, finos o agregados, los investigadores adeptos al análisis de datos individuales dirían que cuando el efecto de la variable independiente es robusto la "inspección visual" de los datos directos apoyaría la conclusión de que el efecto es confiable (e.g., Crosbie, 1999; Reese, 1998). Además, que si se realizara una prueba estadística de contraste sobre los mismos datos, ésta reflejaría exactamente el mismo resultado y por lo tanto sería redundante (Reese, 1998). En cambio, si la variable independiente tiene efectos débiles sobre la variable dependiente, un análisis estadístico podría hacer que el investigador concluyera erróneamente que su variable independiente tiene efectos confiables (Hopkins, Cole, & Mason, 1998; Perone, 1999).

Análisis de Datos Agregados

Algunos psicólogos que se adhieren a las ciencias sociales argumentarían que los datos de una colección de individuos son en realidad la materia prima de la psicología (e.g., "la psicología de las masas") (Brown, 1965). Sin embargo, estos psicólogos serían una minoría en comparación con los que reconocen que la conducta de los organismos individuales es el dato prioritario en la psicología. No obstante, por una razón o por la otra muchos psicólogos adeptos a las ciencias sociales, prefieren caracterizar sus datos como agregados de las conductas individuales, usando estadística descriptiva.

También los partidarios de un enfoque biológico a la psicología como los analistas conductuales pueden emplear estadística para resumir los datos repetidos de sujetos individuales. Si bien usar estadística descriptiva no se encuentra en conflicto con usar observaciones directas de la conducta de los sujetos, tiene la propiedad de invitar inferencias sobre la confiabilidad de los resultados mediante pruebas presumiblemente objetivas, empleando estadística inferencial (e.g., Baron, 1999).

Básicamente la inferencia estadística consiste en expresar la magnitud del efecto de una variable independiente como una proporción del efecto de variables fortuitas y si se trata de comparar la ausencia contra la presencia de una variable independiente, restar la magnitud de una de la otra y expresar esta diferencia como una proporción de la variabilidad fortuita de ambas (Keppel, 1991). Convencionalmente para concluir que el resultado es confiable, la proporción del efecto de la variable independiente debe exceder considerablemente ($p < 0.05$) al efecto de las variables fortuitas (e.g., las "colas" de distribuciones estándar o z).

Críticas al Análisis de Datos Agregados

Como se mencionó antes, la descripción de datos agregados de sujetos individuales no representa problemas particulares para inferir la confiabilidad de un hallazgo científico. Sin embargo, el utilizar pruebas estadísticas para establecer "objetivamente" la confiabilidad es controversial por varias razones.

La primera razón es que las medidas agregadas de sujetos individuales que se incluyen en la prueba estadística no representan a ningún sujeto en particular y por lo tanto es posible que la variable independiente haya tenido efectos diferentes o incluso contrarios a los esperados para algunos sujetos (Perone, 1999; Sidman, 1960). Por esa razón, el resultado de la prueba estadística en la que se basa la inferencia de la confiabilidad del dato no es necesariamente aplicable a todos los sujetos en un grupo (Baron, 1999). En el caso en que la variable dependiente sea la conducta de un mismo sujeto, el problema es el mismo; es decir, que los datos agregados de una condición del estudio pueden no representar la conducta individual en cada repetición de la variable independiente (e.g., Branch, 1999).

La segunda razón por la que las pruebas estadísticas de inferencia son conflictivas, es que su resultado puede depender de detalles de procedimiento. El cómputo de la "significancia" de una prueba estadística incluye como parámetro el número de observaciones en el dato agregado, ya sean sujetos individuales o bien medidas repetidas de un mismo sujeto (Keppel, 1991).

Como regla general, a mayor cantidad de datos en las medidas agregadas, la "significancia" de la prueba también aumenta (e.g., Hopkins et al., 1998; Perone, 1999). Este detalle de procedimiento ha tenido como consecuencia el que ocasionalmente se reporten resultados que no son robustos o sistemáticos

Por esta razón, la inferencia sobre la confiabilidad de un hallazgo científico puede depender más del propio investigador que en la genuina diferencia entre las medidas agregadas.

La tercera razón por la que la inferencia sobre el efecto de la variable independiente basada en pruebas estadísticas es controversial radica en que la replicabilidad del hallazgo de una investigación en estudios idénticos subsecuentes se establece en términos de la probabilidad de que el efecto de la variable independiente exceda al efecto de las variables fortuitas (i.e., $p < .05$, conocida como “probabilidad de replicación”) (e.g., Killeen, 2005, 2006; Miller, 2009; Sanabria & Killeen, 2007). En consecuencia, la replicación de un hallazgo puede depender de factores fortuitos más que en el genuino efecto de la variable independiente.

La cuarta razón por la que la estadística inferencial es controversial radica en su supuesta objetividad para establecer la confiabilidad de los hallazgos. La elección del análisis estadístico apropiado para una serie de datos agregados debe adherirse a una serie de reglas específicas establecidas (e.g., supuestos que los datos deben cumplir, número de observaciones realizadas) (Tabachnick & Fidell, 1989). No obstante, en algunos estudios se ha documentado que la interpretación de las reglas depende del investigador (Perone, 1999). Aún más, una serie de datos agregados, ya sean de repeticiones de observaciones de un mismo sujeto o de una colectividad de sujetos, pueden analizarse utilizando diferentes pruebas estadísticas y los investigadores no siempre coinciden sobre el análisis pertinente para establecer la confiabilidad de los datos (e.g., Nourbakhsh & Ottenbacher, 1994). Aún más, se ha demostrado que aún cuando se cumplan los supuestos en los que se basan los análisis estadísticos, la obtención de un resultado significativo variará dependiendo del análisis específico utilizado (e.g., Nourbakhsh & Ottenbacher, 1994; Tabachnick & Fidell, 1989).

Defensa del Análisis de Datos Agregados

Respecto a la crítica de la representación de sujetos individuales en la medida agregada, los investigadores adeptos a la estadística responderían que debido a variables fortuitas las diferencias individuales son inevitables y que por esta razón es conveniente dejar que tales diferencias se cancelen mutuamente en la medida agregada (Crosbie, 1999; Reese, 1998). El mismo argumento aplica al caso de las medidas repetidas de un mismo sujeto. Es esperable que en cada repetición de la variable independiente estén presentes diferentes condiciones fortuitas, que al pasar inadvertidas alterarán la conducta de cada sujeto en la medida agregada de manera diferente (Reese, 1999). En breve, que la variabilidad en el dato es un hecho inevitable en la investigación y que la medida agregada tiene la virtud de “eliminar” tal variabilidad, enfatizando comunidades (Reese, 1998; Shull, 1999). Por

ejemplo, la ejecución particular de un sujeto puede depender de una situación particular y no reflejar generalidades en la conducta. En cambio, ciertos principios o leyes basados en datos agregados reflejan la generalidad de un hallazgo como en el caso del aumento suave de un gradiente de respuesta, de discriminación o generalización, que aunque están basados en datos agregados expresan características fundamentales e importantes del control temporal de la conducta o del control que ejerce un estímulo (Shull, 1999).

Con respecto a la crítica sobre la manipulación de la significancia de una prueba estadística por parte del investigador, los investigadores adeptos a la estadística responderían que desde el diseño del estudio calcularon cuidadosamente el número de sujetos o de datos de un mismo sujeto que posteriormente incluyeron en su prueba. La forma como se calcula el número de datos que participan en las medidas agregadas se conoce como pruebas de poder, que en breve determinan el número óptimo de datos agregados que eviten que la prueba estadística subestime o sobreestime la magnitud del efecto de la variable independiente (Keppel, 1991).

Respecto a la crítica de que la replicación de un hallazgo depende de factores fortuitos y no necesariamente de la magnitud del efecto de la variable independiente, los investigadores que favorecen el uso de estadística argumentarían que aún en la observación de la conducta individual bajo situaciones controladas, existen variables fortuitas como cambios en el medio ambiente o en el millieu interno de los sujetos, por ejemplo, variabilidad debida a diferencias individuales o intrínseca (Sidman, 1960) que afectan su ejecución y pueden afectar la replicabilidad de un hallazgo (Reese, 1998). También argumentarían que aunque todos los investigadores deben lidiar con los efectos del azar producto de variables que no están bajo el control del propio investigador, ya sea de datos de sujetos individuales o de datos agregados, la estadística ofrece la ventaja sobre la "inspección visual" de los datos de poder computar el efecto de tales efectos fortuitos (e.g., Baer, 1977). El uso de estadística inferencial para determinar la magnitud del efecto de una variable independiente, aislando los efectos azarosos de variables individuales, no representa entonces ningún problema mientras se utilice correctamente y por las razones apropiadas (Shull, 1999).

Relativo a la crítica de la supuesta "objetividad" del análisis estadístico inferencial, los investigadores que favorecen el uso de éste argumentarían que la estadística es sólo una herramienta que ayuda al investigador a tomar decisiones sobre la confiabilidad de sus datos, pero no remueve su responsabilidad para decidir sobre la utilidad o generalidad de sus hallazgos (e.g., Crosbie, 1999; Reese, 1998, 1999; Shull, 1999). También dirían que, a diferencia de la "inspección visual" de los datos para la que no hay reglas establecidas (Perone, 1999; Sidman, 1960), la estadística es una estrategia probada y razonable para tomar decisiones sobre la confiabilidad de los datos

para la que no hay reglas establecidas (Perone, 1999; Sidman, 1960), la estadística es una estrategia probada y razonable para tomar decisiones sobre la confiabilidad de los datos (e.g., Michael, 1974). De hecho, los resultados de investigaciones que incluyen inferencias estadísticas sobre la confiabilidad de datos agregados han mostrado efectos robustos y consistentes de variables independientes que afectan tanto a individuos aislados como a colectividades de individuos (Shull, 1999).

Un punto de vista razonado

El requisito básico de cualquier estudio psicológico consiste en comparar la variable dependiente en ausencia o en presencia de una variable independiente. Aún en el caso de los llamados estudios correlacionales, la intención del investigador es mostrar la covariación entre dos series de datos, ya sea que se considere a una de las dos como la variable independiente o bien ambas series de variables se consideren la consecuencia de una tercera. En este último caso, el estudio psicológico demanda un punto de comparación (e.g., un caso a donde no existe tal correlación). El grado de asociación de una variable independiente con una dependiente o bien entre dos variables dependientes se conoce como confiabilidad del hallazgo científico (Pultchik, 1968).

Como se discutió anteriormente, existen por lo menos dos tradiciones para argumentar la confiabilidad de un hallazgo, la originaria de las ciencias biológicas y la originaria de las ciencias sociales. Mientras que la primera privilegia el análisis directo de datos individuales, la segunda privilegia el análisis estadístico de datos agregados. En las secciones anteriores de este trabajo se presentaron las premisas básicas de ambas estrategias, así como las críticas que hacen los partidarios de un enfoque o del otro. También se presentaron las respectivas defensas de los partidarios de ambos enfoques. La conclusión de esta discusión puede resumirse diciendo que los dos enfoques tienen fortalezas y debilidades.

Es común que los partidarios de alguno de los dos enfoques coordinen el análisis de los datos individuales o agregados con algún tipo de diseño experimental. Sin embargo, no parece existir una razón de fondo para privilegiar la inspección directa de las medidas repetidas producto de cualquier diseño experimental sobre el análisis de datos agregados de la conducta individual. Es cierto que las disciplinas que se derivan de las ciencias biológicas involucran la repetición de observaciones sobre la conducta de sujetos individuales. Sin embargo, también es cierto que las disciplinas que se derivan de las ciencias sociales involucran la repetición de observaciones de una colección de sujetos. Si es cierto que el análisis de datos individuales o de datos agregados es independiente del diseño del estudio, entonces representar los datos individuales directamente o de forma agregada depende

en gran medida de la preferencia del investigador.

La razón de fondo para preferir el análisis directo de los datos individuales o el análisis estadístico de datos agregados, se reduce a la naturaleza de la pregunta de investigación. El avance en el conocimiento científico no ocurre en un vacío de información. Por el contrario, cada pregunta de investigación se deriva lógicamente de la investigación anterior. Tal investigación anterior puede provenir de cualquiera de las dos disciplinas, la biológica o las ciencias sociales, y dependiendo de las tradiciones y convenciones de cada una la pregunta de investigación involucra uno de los dos tipos de análisis de los datos. Por ejemplo, en la tradición analítico-conductual los antecedentes de una pregunta de investigación seguramente involucran el análisis de medidas repetidas de la conducta de pocos sujetos. Por lo tanto, si la pregunta de investigación que se pretende responder representa una extensión de la investigación anterior, lógicamente requiere el examen de medidas repetidas de la conducta de unos cuantos sujetos.

Otro ejemplo, en la tradición de alguna disciplina derivada de las ciencias sociales como podría ser la psicología social, es muy probable que la investigación anterior involucre el examen de datos agregados de la conducta de los participantes. Es muy probable que se utilicen promedios y medidas de dispersión como estadística descriptiva. También es muy probable que se utilicen medidas de asociación o diferencia para establecer la confiabilidad de los hallazgos. También es muy probable que los grados de asociación o diferencia entre los datos agregados se sometan a estadística inferencial para decidir si son o no confiables. Si la pregunta de investigación que se pretende responder representa una extensión de la investigación anterior, lógicamente requiere el análisis estadístico de los datos tanto a nivel descriptivo como inferencial en los mismos términos que los antecedentes del estudio. Esto no excluye desde luego que con el fin de responder sin ambigüedad la pregunta de investigación, el investigador decida analizar sus datos, ya sea de forma individual o agregada, en adición a los análisis que directamente le permiten la comparación con los de investigación anterior.

Desde luego que los partidarios del enfoque alternativo criticarán el análisis de los datos del estudio en cuestión, pero esto no será razón para que el investigador modifique el análisis de sus datos, si no por otra razón porque en conformidad con las tradiciones de su campo las revistas especializadas lo forzarán a usar un tipo particular de análisis. Después de todo, hay que reconocer que la investigación científica sigue las convenciones que dicta la comunidad científica y que es en este contexto en el que se desarrolla cualquier avance del conocimiento (Kuhn, 1962).

Como ya se mencionó, en la tradición analítico-conductual se privilegia el análisis directo de datos individuales (Sidman, 1960; Skinner, 1938). Conforme esta tradición, la mayoría de los artículos publicados en revistas especializadas como la RMAC o el JEAB o bien no han incluido ningún tipo de estadística o bien sólo han empleado estadística descriptiva con el fin de sumarizar los datos individuales (ver Figura 1).

Es importante mencionar, no obstante, que aún en el caso de los artículos en los que se presentan datos agregados, es común que también se incluyan figuras o tablas con datos individuales. Si bien el uso de estadística descriptiva es perfectamente compatible con la tradición analítico-conductual, es un misterio la razón por la que ha aumentado el número de artículos publicados en la RMAC y en el JEAB en los que se ha utilizado estadística inferencial para determinar la confiabilidad de los hallazgos. El uso de estadística inferencial para establecer la confiabilidad de los hallazgos en los artículos publicados en la RMAC y en el JEAB ha ido en continuo aumento desde los 80's y hasta la fecha. Este aumento ha sido tal que en la RMAC a partir del 2000 y en el JEAB a partir de 1993 la proporción de artículos que incluyeron estadística inferencial (alrededor de un 50%) excedió a la proporción de artículos en los que se empleó estadística descriptiva (alrededor de un 30%) y sólo en el 20% se utilizó un análisis directo de datos individuales (ver Figura 1).

Este aumento en el uso de análisis estadísticos inferenciales para establecer la confiabilidad de los hallazgos en revistas especializadas es un misterio si se considera que después de todo en la contraportada del JEAB se especifica que dicha revista es primordialmente para publicar experimentos originales relevantes a la conducta de organismos individuales.

Ciertamente, ni en la RMAC ni en el JEAB se exige el uso de estadística inferencial para especificar la confiabilidad de un hallazgo científico. Si bien Baum (2002) argumentó un cambio de paradigma en el análisis de la conducta, del molecular en el que se consideraba a la conducta y a los estímulos medio-ambientales como eventos discretos, al molar, en el que se considera a la conducta como agregados de patrones de acciones que ocurren en un determinado tiempo, este cambio paradigmático sólo debería haber repercutido en un mayor uso de estadística descriptiva, pero no permite explicar el aumento en el uso de estadística inferencial por parte de los analistas conductuales.

Es posible que, como se dijo antes, el uso de estadística descriptiva haya favorecido la prueba de la confiabilidad de los datos mediante estadística inferencial, especialmente en aquellos casos en los que existe mucha variabilidad en los datos individuales y la "inspección visual" no sea confiable. Otra posibilidad, es que para responder su pregunta de investigación sin ambigüedad o para poder comparar sus datos con los de investigación anterior los analistas conductuales hayan considerado necesaria o útil la inferencia de

la confiabilidad de sus hallazgos mediante estadística. Otra posibilidad es que, dado que en la mayoría de las otras áreas de la psicología se utiliza estadística inferencial, los analistas conductuales estén adhiriéndose cada vez más a las convenciones que dicta la mayoría de la comunidad científica psicológica.

Referencias

- Baer, D. M. (1977). Perhaps it would be better not to know everything. *Journal of Applied Behavior Analysis*, 10, 167-172.
- Baron, A. (1999). Statistical inference in behavior analysis: Friend or foe? *The Behavior Analyst*, 22, 83-85.
- Baum, W. M. (2002). From molecular to molar: A paradigm shift in behavior analysis. *Journal of the Experimental Analysis of Behavior*, 78, 95-116.
- Danov, S. E., & Symons, F. J. (2008). A survey evaluation of the reliability of visual inspection and functional analysis graphs. *Behavior Modification*, 32, 828-839.
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. *Journal of Applied Behavior Analysis*, 12, 573-579.
- Fish, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes*, 54, 137-154.
- Francis, N. A. (2005). Single subject trials in primary care. *Postgraduate Medicine Journal*, 81, 547-548.
- Hopkins, B. L., Cole, B. L., & Mason, T. L. (1998). A critique of the usefulness of inferential statistics in applied behavior analysis. *The Behavior Analyst*, 21, 125-137.
- Kazdin, A. E. (1982). Single-case research designs: *Methods for clinical and applied settings*. Nueva York, E. U.: Oxford University Press.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Nueva Jersey, E. U.: Prentice Hall.
- Killeen, P. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, 16, 345-353.
- Killeen, P. (2006). Beyond statistical inference: A decision theory for science. *Psychonomic Bulletin & Review*, 13, 549-562.
- Kuhn, T. S. (1962). *The structure of scientific revolution*. Chicago, E. U.: University of Chicago Press.
- Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis*, 7, 647-653.
- Miller, J. (2009). What is the probability of replicating a statistically significant effect? *Psychonomic Bulletin & Review*, 16, 617-640.

- Newcomb, R. G. (2005). Should the single subject design be regarded as a valid alternative to the randomized controlled trial? *Postgraduate Medicine Journal*, 81, 546-547.
- Nourbakhsh, M. R., & Ottenbacher, K. J. (1994). *The statistical analysis of single-subject data: A comparative examination*. Recuperado el 2 de abril de 2010 de . . .
http://findarticles.com/p/articles/mi/_hb237/is_n8_v74/ai_n28643432/
- Perone, M. (1999). Statistical inference in behavior analysis: Experimental control is better. *The Behavior Analyst*, 22, 109-116.
- Plutchik, R. (1968). *Foundations of experimental research*. Nueva York, E. U.: Harper's Experimental Psychology Series.
- Reese, H. W. (1998). Utility of group methodology in behavior analysis and developmental psychology. *Revista Mexicana de Análisis de la Conducta*, 24, 137-151.
- Reese, H. W. (1999). Problems of statistical inference. *Revista Mexicana de Análisis de la Conducta*, 25, 39-68.
- Sanabria, F., & Killeen, P. (2007). Better statistics for better decisions rejecting hypotheses statistical tests in favor of replication statistics. *Psychology in the Schools*, 44, 471-481.
- Shull, R. L. (1999). Statistical inference in behavior analysis: Discussant's remarks. *The Behavior Analyst*, 22, 117-121.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Nueva York, E. U.: Basic Books.
- Skinner, B. F. (1938). *The behavior of organisms*. Nueva York, E. U.: Appleton-Century.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper, Collins Publishers, Inc.