

Sobre las hipótesis direccionales y no direccionales

©Pedro Morales

• Universidad Pontificia Comillas • Madrid •
Facultad de Ciencias Humanas y Sociales
(Última revisión: 7 de Febrero de 2013)

Índice

1. ¿Qué es una hipótesis?	3
2. ¿Es posible investigar sin hipótesis? Los estudios exploratorios	3
3. Hipótesis Nula e Hipótesis Alterna	4
4. Hipótesis direccionales e hipótesis no direccionales	5
5. Consecuencias de escoger hipótesis direccionales	8
6. Cuándo se pueden utilizar hipótesis direccionales	9
7. Qué vemos en la práctica	9
8. Problemas con las hipótesis direccionales	10
8.1. Problemas de interpretación cuando los resultados son estadísticamente significativos pero en la dirección opuesta a la especificada	10
8.2. Las Hipótesis direccionales y el descubrimiento por sorpresa	11
9. Qué dicen los diversos autores	12
10. Qué hacer en la práctica	13
10.1. Preferencia por las hipótesis no direccionales	13
10.2. No limitarse a calcular e interpretar, como último dato, los valores relacionados con la significación estadística (t , F , p)	15
11. Referencias bibliográficas	17

1 ¿Qué es una hipótesis?

Si podemos decir que en términos generales *investigar es responder a preguntas*, una hipótesis es una *respuesta tentativa* a nuestra pregunta; es una respuesta que queremos confirmar. Una hipótesis es una *predicción*, una afirmación que describe en términos específicos *el resultado esperado* de nuestra investigación.

Si en una investigación se parte de hipótesis que se quieren confirmar, estas hipótesis *hay que justificarlas previamente* de alguna manera (normalmente en función de alguna teoría, o de datos previos o de nuestra propia observación e intuición). Aunque no es raro encontrar como punto de partida de una investigación hipótesis sin ningún tipo de justificación; en principio esta práctica es incorrecta; debemos tener *algún tipo* de justificación para proponer una hipótesis. Un mismo estudio puede tener una o varias hipótesis y también puede haber estudios sin hipótesis formales previamente formuladas, como sucede en los estudios meramente exploratorios.

2. ¿Es posible investigar sin hipótesis? Los estudios exploratorios

Sí se puede plantear una investigación sin hipótesis previas; *not all studies have hypotheses* (Trochim, 2006)¹; *no todas las investigaciones cuantitativas plantean hipótesis* (Hernández Sampieri, Fernández Collado, y Baptista Lucio (2008:122).

El comenzar una investigación proponiendo hipótesis:

1º No hay que entenderlo como un *requisito necesariamente obligatorio* de manera que una investigación o una tesis no son académicamente correctas sin hipótesis previas,

2º Este modelo (*teoría-hipótesis-confirmación*) corresponde al paradigma *deductivo* de investigación, pero no es el único paradigma o modelo posible.

También cabe el paradigma *inductivo* en el que a partir de unos resultados se proponen (como conclusión, no como punto de partida) unas hipótesis que se pueden confirmar en otros estudios posteriores. Estos dos modelos o paradigmas, deductivo e inductivo, están representados en la figura 1²

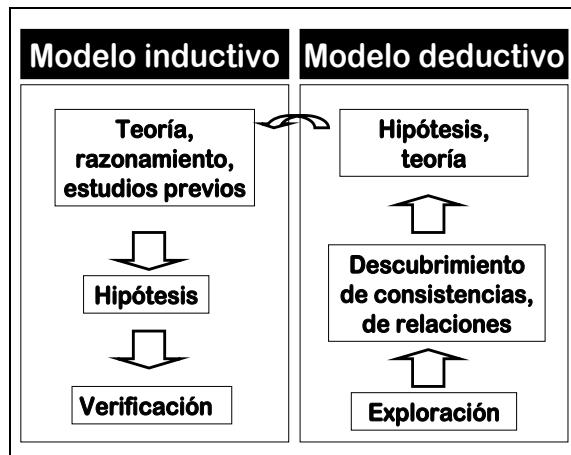


Figura 1

En estos estudios de carácter *exploratorio* (que también suelen denominarse en inglés *data mining*) se pretende verificar si existen relaciones sistemáticas (*patterns*) entre variables (o

¹ La última revisión consultada de esta sección de Trochim, (*hypothesis*) es de 10/20/2006.

² Adaptada muy libremente de Trochim (2006)

diferencias entre grupos, etc.) cuando no hay (o *no están claras*) hipótesis *a priori*³. Estos estudios exploratorios sí suelen tener un objetivo, un interés en un determinado ámbito, un sentido de dirección.

Son dos modelos que no se excluyen necesariamente; de las conclusiones de un estudio exploratorio (*paradigma inductivo*) pueden derivarse hipótesis que pueden ser el punto de partida de otra investigación (*paradigma deductivo*) y en una misma investigación puede haber tanto hipótesis previas como una *exploración* de los datos sin hipótesis previas. A veces lo que tenemos son *sospechas* que no proponemos formalmente como hipótesis porque nos falta una justificación suficiente y preferimos proponer nuestra investigación como una exploración. Además siempre cabe el *descubrimiento por sorpresa* (así descubrió Fleming la penicilina).

Por otra parte en estudios básicamente exploratorios (sin hipótesis claras formuladas y justificadas previamente) es normal y correcto al evaluar los resultados utilizar expresiones del estilo *Hipótesis Nula rechazada* o *Hipótesis Alterna aceptada*, aunque no estaba prevista ninguna hipótesis. La Hipótesis Nula, formulada o no formulada previamente, está implícita en el mismo paradigma y procedimiento para llevar a cabo un contraste de medias y otros análisis semejantes (simplemente determinamos la probabilidad de obtener un determinado resultado). También se dan investigaciones en las que se pretenden ambas finalidades, confirmar algunas hipótesis y a la vez explorar los datos disponibles.

3. Hipótesis Nula e Hipótesis Alterna

Repasamos la distinción entre Hipótesis Nula e Hipótesis Alterna. Cuando decimos que tenemos *una hipótesis*, realmente estamos *pensando* en dos hipótesis o respuestas posibles:

a) Una hipótesis (o resultado *esperado*) describe *nuestra predicción*; es la Hipótesis Alterna que podemos expresar de diversas maneras, *estas dos medias son distintas, su diferencia es significativamente mayor que la diferencia cero, las dos medias son de muestras que proceden ‘de poblaciones con distinta media’*, etc.

b) La otra hipótesis describe otros posibles resultados que niegan nuestra predicción; es la Hipótesis Nula, *estas dos medias pertenecen a muestras ‘de la misma población’ con una única media, sin más diferencias entre las medias que las puramente aleatorias; es una diferencia que pertenece a una ‘población de diferencias’ cuya media es cero*.

La hipótesis Alterna y la Hipótesis Nula se excluyen mutuamente. En términos formales solemos formular las dos hipótesis, la Hipótesis Alterna (la *nuestra*, la que queremos confirmar) y también la Hipótesis Nula (la que niega nuestra hipótesis).

Aunque no formulemos explícitamente la Hipótesis Nula (no siempre se formula de manera rutinaria), *siempre está implícita* y esta Hipótesis Nula la que *en primer lugar ponemos a prueba* mediante las pruebas o tests estadísticos. En las *pruebas de significación* lo que hacemos es examinar la *viabilidad* de la Hipótesis Nula, que será aceptada o no aceptada a la luz de los datos disponibles en nuestro estudio. Si queremos probar que nuestra hipótesis (Hipótesis Alterna) es verdadera, en primer lugar asumimos que nuestra hipótesis *no* es verdadera (es lo que decimos al formular la Hipótesis Nula), y es esta suposición la que ponemos a prueba.

En muchos estudios es frecuente formular (redactar, manifestar) solamente las hipótesis que queremos confirmar (*hipótesis de investigación*, las Hipótesis Alternas), y mencionar la Hipótesis Nula solamente al final de la prueba estadística, cuando se rechaza o no se rechaza; en

³ “Exploratory Data Analysis as opposed to traditional *hypothesis testing* designed to verify *a priori* hypotheses about relations between variables” (StatSoft, 2007, *online*; ver en el índice: *Data Mining Techniques, Exploratory Data Analysis (EDA) and Data Mining Techniques*).

otros no se menciona la Hipótesis Nula en ningún momento⁴. Si nuestra hipótesis (Hipótesis Alterna) no se ve apoyada por los resultados obtenidos, es porque aceptamos (con más propiedad *no rechazamos*) la Hipótesis Nula⁵.

En ocasiones nuestra predicción (nuestra hipótesis) es que *no hay diferencia* (o no hay cambio, no hay relación), y en estos casos lo que buscamos es confirmar la Hipótesis Nula.

4. Hipótesis direccionales e Hipótesis no direccionales

Vamos a tratar de manera más específica sobre las hipótesis *direccionales* o *no direccionales* ¿Por qué dedicamos una reflexión especial a algo que podemos encontrar en todos los textos de estadística inferencial? Como iremos comentando, el tema de las hipótesis *direccionales* o *no direccionales* no ha estado exento de polémica, y con frecuencia causa cierta *perplejidad*⁶.

Todos recordamos sin mucho esfuerzo el valor de $z = 1.96$ en la *distribución normal*: cuando superamos este valor en un contraste de medias es cuando decimos que se trata de una *diferencia estadísticamente significativa*. Aunque hacemos esta exposición pensando en el caso frecuente del *contraste de medias*, es igualmente válida para otros planteamientos en los que buscamos un resultado *estadísticamente significativo*.

En el numerador de la *t de Student* tenemos *una diferencia entre dos diferencias*: una diferencia (la de las dos medias que comparamos, $\bar{X}_1 - \bar{X}_2$) y una *diferencia media de cero* (la que encontraríamos si no hubiera más diferencias que las puramente *casuales*, o explicables por el *error muestral*)⁷. La probabilidad de que la diferencia entre las dos medias ($\bar{X}_1 - \bar{X}_2$) supere a la *diferencia media de cero* en 1.96 *errores típicos* es del 2.5 %, y a la inversa tenemos lo mismo; la probabilidad de que la segunda media (\bar{X}_2) sea mayor es también del 2.5%.

El utilizar el valor de 1.96 (en muestras grandes) como referencia para no aceptar la Hipótesis Nula supone que nuestra hipótesis es *bidireccional*, tal como aparece en la figura 1. Es decir, la probabilidad de que una diferencia supere el valor de 1.96, si incluimos las dos direcciones, es del 5% (2.5% + 2.5%), que corresponde al *nivel de confianza* habitual (.05). Por eso recordamos con facilidad el valor de 1.96, porque la práctica habitual es utilizar hipótesis bidireccionales.

⁴ No hay una práctica generalizada ni un consenso entre investigadores sobre qué hipótesis se deben explicitar en una investigación; no hay normas establecidas (Hernández Sampieri, Fernández Collado, y Baptista Lucio, 2002:165); más bien suele haber normas *situacionales* (por ejemplo las que da *un profesor* a sus alumnos, o las que da *una revista* a sus colaboradores). Cuando es el investigador quien decide, debe pensar en los receptores de la investigación (los lectores pueden no entender qué es la Hipótesis Nula). Como orientación general pensamos que al menos deben formularse las hipótesis de investigación (*alternas*) si las hay y pueden justificarse.

⁵ *Aceptar* no es lo mismo que *no rechazar*; son expresiones que pueden prestarse a confusión y conviene entender lo que se quiere decir. En principio es incorrecta la expresión *Hipótesis Nula aceptada* si entendemos *aceptada* como sinónimo de *verdadera* (confusión relativamente frecuente); nunca probamos que la Hipótesis nula es verdadera, simplemente fracasamos en nuestro intento de rechazarla; la Hipótesis Nula *se rechaza o no se rechaza*.

⁶ Hablando con propiedad las *hipótesis* son *o direccionales* (señalamos la dirección de la diferencia) o *no direccionales* (no señalamos la dirección de la diferencia). Las expresiones comunes *una cola* o *dos colas*, *unilaterales* o *bilaterales*, etc., se aplican a las distribuciones y a los *tests o pruebas estadísticas*, no a las hipótesis (Toothaker y Miller, 1996: 323).

⁷ *Causalidad*, *azar*, etc. son expresiones habituales en el lenguaje informal y fácilmente comprensibles; realmente lo que queremos decir con estos términos es que la diferencia se explica suficientemente por el *error muestral*, y por *error muestral* (en la terminología académica) entendemos la *variabilidad normal* que se da en cualquier conjunto de datos u observaciones; aquí naturalmente la *normalidad* es estadística; es normal lo que es habitual o frecuente. Los límites de esta normalidad los establecemos con el *nivel de confianza* (entre qué límites máximo y mínimo *estimamos* que se encuentra lo que denominamos *normal*).

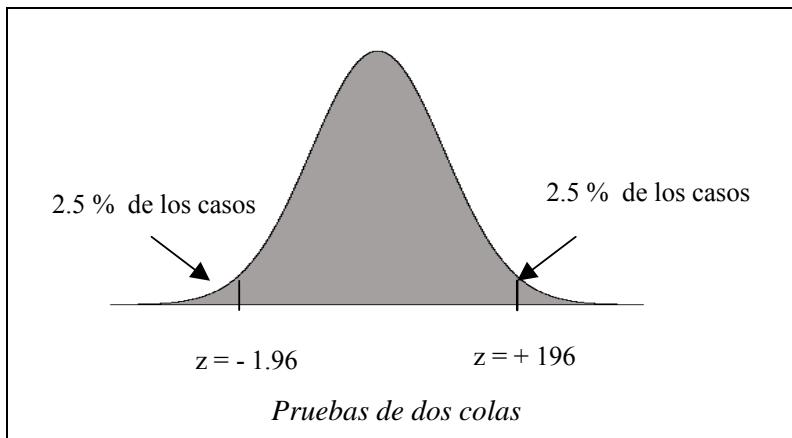


Figura 1

En realidad muchas de nuestras hipótesis tal como las concebimos, o quizás la mayoría, son *unidireccionales* (*este grupo* tiene una media mayor que el otro, ha habido un cambio pero en *esta dirección*, etc.). Con una hipótesis *unidireccional* el valor que tendríamos que recordar con facilidad es 1.64, y no 1.96, porque la probabilidad de obtener un valor de t (o z) superior a 1.64 es del 5% tal como se ve en la figura 2.

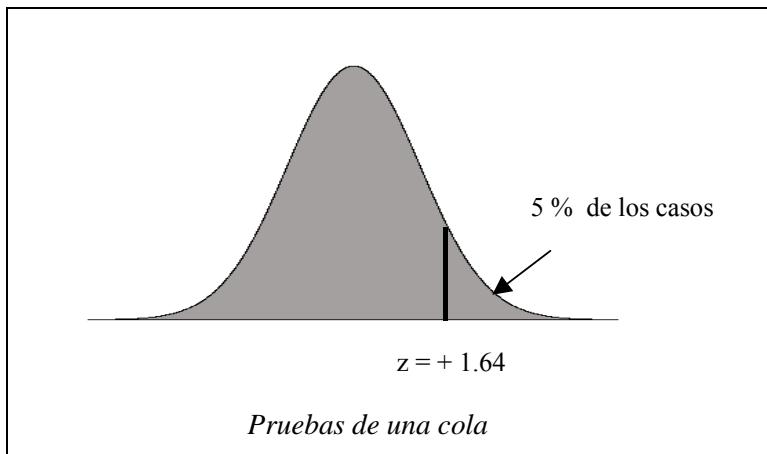


Figura 2

Esta aparente y frecuente inconsistencia entre la teoría (*proponer o suponer hipótesis unidireccionales*) y lo que vemos que se hace habitualmente (*utilizar los valores propios de las hipótesis bidireccionales*) es lo que queremos examinar más despacio, paso a paso.

Mohr (1992), en una conocida obra de alta divulgación sobre *significance testing*, reconoce que la mayor parte de las hipótesis son (o deberían ser con frecuencia) *unidireccionales*, pero que en la práctica es habitual utilizar los valores de t de las pruebas bidireccionales, con lo que el nivel usual de $\alpha = .05$ es realmente de .025 y es lo que debería manifestarse si hay una hipótesis clara unidireccional previa.

Sí parece cierto, como ya hemos dicho, que lo que se ve con mucha frecuencia es utilizar los valores propios de las pruebas bidireccionales cuando las hipótesis especificadas, a veces de manera implícita, son unidireccionales.

La *Hipótesis Nula* se establece en términos de una igualdad: $\mu_1 = \mu_2$; las dos muestras proceden de la misma población con una única media.

La *Hipótesis Alterna* se establece en términos de una *desigualdad* que admite dos modalidades:

- a) *Hipótesis Alterna* en la que se especifica la dirección de la diferencia: $\mu_1 > \mu_2$, o $\mu_2 > \mu_1$ (denominadas hipótesis, o pruebas, *unidireccionales*, *unilaterales* o de *una cola*; *one-tailed* en inglés).
- b) *Hipótesis Alterna* en la que no se especifica la dirección de la diferencia: $\mu_1 \neq \mu_2$ (denominadas también pruebas, *bilaterales* o de *dos colas*, *two-tailed* en inglés, y con más propiedad *no direccionales*,).

Podemos establecer por lo tanto tres hipótesis alternas (sin especificar la dirección o especificando la dirección en uno u otro sentido), con sus correspondientes hipótesis nulas.

La Hipótesis Nula es la *negación* de la Hipótesis Alterna; si aceptamos la Hipótesis Nula como verdadera (o con más propiedad, como *no falsa*), no aceptaremos la Hipótesis Alterna porque la juzgamos como *no probada*, y viceversa, si no aceptamos la Hipótesis Nula, aceptaremos la Hipótesis Alterna.

Conviene recordar que las expresiones comunes *aceptar* la Hipótesis Nula (porque es verdadera) o *rechazar* la Hipótesis Nula (porque es falsa) no son las idóneas a pesar de su uso frecuente; la Hipótesis Nula *la rechazamos o no la rechazamos*; en rigor nunca probamos que la Hipótesis Nula es verdadera. Lo único que hacemos es determinar la probabilidad de que se dé una diferencia de una determinada magnitud en el caso de que la diferencia se deba a factores puramente casuales o aleatorios.

No probar que una proposición es verdadera no es lo mismo que probar que dicha proposición es falsa, y esta distinción (que lleva a muchos errores de interpretación cuando no se entiende bien) es importante en este contexto, por eso, y hablando con propiedad, *nunca probamos que la Hipótesis Nula es verdadera*; simplemente fracasamos en nuestro intento de rechazarla, de poder afirmar que es falsa. Cualesquiera que sean las expresiones que utilizamos, hay que entender lo que quieren decir.

Puede merecer la pena ilustrar esto con un ejemplo sencillo.

1. Supongamos que nuestra hipótesis (alterna, la hipótesis que quiere probar el investigador) es ésta: *a los italianos les gusta más la ópera que a los españoles*.
2. Comparamos dos muestras (de 10 sujetos cada una), una de italianos y otra de españoles (igualados en nivel cultural, etc.) y vemos que la media de los italianos es mayor, pero la t de Student queda lejos del valor crítico (por ejemplo $t = 1.89$ en vez de $t = 2.1$ que vemos en las tablas para ese número de sujetos); esa diferencia no se aparta significativamente de una media de cero.
3. ¿Hemos probado que a los italianos les gusta más la ópera que a los españoles? No lo hemos probado (no podemos *extrapolar* una diferencia distinta de cero a otros posibles pares de muestras).
4. ¿Hemos probado que *no hay diferencia* entre españoles e italianos? *Tampoco* lo hemos probado (quizás con una muestra mayor podríamos ver con mucha seguridad una diferencia que no vemos en muestras muy pequeñas).

Aceptar la Hipótesis Nula no significa por lo tanto que hemos probado la hipótesis contraria a la Hipótesis Alterna, sino simplemente que no hemos confirmado nuestra Hipótesis (la Hipótesis Alterna propuesta, la del investigador). Esto es importante porque, volvemos a insistir, no es lo mismo *no probar* que una diferencia es verdadera que *probar que no* es verdadera (en

otra ocasión, con un mejor diseño, con una muestra mayor, etc., podremos quizás rechazar la Hipótesis Nula y aceptar la Hipótesis Alterna que afirma la diferencia).

Estas hipótesis suelen simbolizarse tal como aparece en la figura 4 (μ es el símbolo de la media de una población)⁸.

	Hipótesis Alterna (H_1)	Hipótesis Nula (H_0)
Hipótesis no direccionales (o bidireccionales) Establecemos en la H_1 que las medias son distintas, pero no predicimos de antemano cuál de las dos es mayor	A: $\mu_1 \neq \mu_2$ $\mu_1 - \mu_2 \neq 0$	$\mu_1 - \mu_2 = 0$ $\mu_1 = \mu_2$
Hipótesis direccionales En H_1 establecemos que las medias son distintas y además especificamos de antemano cuál de las dos esperamos que sea mayor	B: $\mu_1 > \mu_2$	$\mu_1 < \mu_2$ $\mu_1 = \mu_2$
	C: $\mu_1 < \mu_2$	$\mu_1 > \mu_2$ $\mu_1 = \mu_2$

Figura 4

1º En la hipótesis A de la figura 4 *no especificamos la dirección de la diferencia*:

Si *una de las dos medias* es mayor que la otra rechazaremos la Hipótesis Nula y podremos aceptar la Hipótesis Alterna, diremos que sí hay diferencia (afirmaremos que *las muestras proceden de poblaciones distintas con distinta media*) y esto tanto si $\bar{X}_1 > \bar{X}_2$ como si $\bar{X}_1 < \bar{X}_2$; se trata de un contraste *bilateral* o de *dos colas*.

2º En las Hipótesis B y C *sí indicamos la dirección de la diferencia*:

No decimos solamente que las medias son distintas (que la diferencia entre las poblaciones *no* es 0), sino que una media determinada es mayor que la otra; se trata en este caso de hipótesis *unilaterales* o *unidireccionales*.

Aceptaremos la Hipótesis Nula tanto si no hay diferencia entre las medias como si la media hipotéticamente menor es, según vemos en las tablas de la distribución normal, significativamente mayor (es decir, no probamos la Hipótesis que contradice la nuestra aunque parezca probada... consideramos este resultado como dentro de lo aleatorio y esto, como veremos, presenta problemas).

5. Consecuencias de escoger hipótesis direccionales

Nos fijamos en dos consecuencias de muy diverso orden: si la hipótesis es direccional (o unilateral):

1º En primer lugar si la hipótesis es unidireccional, la Hipótesis Nula se rechaza (*no se acepta*) con más facilidad porque necesitamos un valor menor de t ; con muestras grandes (con un nivel de confianza del 95%) nos basta $z = 1.64$ y no $z = .196$.

2º En segundo lugar la investigación tiene más probabilidades de ser publicada (al menos satisface más al investigador) si obtenemos resultados *estadísticamente significativos*.

El escoger una hipótesis *direccional* o *no direccional* (o utilizar pruebas de *una o dos colas*) afecta a los resultados de la investigación: si la hipótesis es *unilateral*, se rechaza la Hipótesis Nula con un *valor menor de t (o z)*; en este caso sólo tenemos en cuenta las probabilidades

⁸ Las hipótesis expresadas con estos símbolos se denominan *hipótesis estadísticas* (Hernández Sampieri, Fernández Collado, y Baptista Lucio, 2002:161).

asociadas a uno de los extremos (a una de las dos *colas*) de la distribución normal (tabla 1, valores para muestras grandes):

Se rechaza (no se acepta) la Hipótesis Nula con una probabilidad de error de:		
Con un valor de t (o z) de	Si la hipótesis es direccional (prueba de una cola)	Si la hipótesis es bilateral (prueba de dos colas)
1.64	.05	.10
1.96	.025	.05
2.32	.001	.02
2.57	.005	.01
3.30	.0005	.001

Tabla 1

Consecuentemente la *significación estadística* se consigue más fácilmente con *hipótesis unidireccionales*.

Para un mismo valor de t (ó z), en las hipótesis bilaterales las probabilidades de error al rechazar la hipótesis nula son el doble que en las hipótesis de una cola. Dicho en otros términos, las pruebas estadísticas tienen más *potencia* cuando son unilaterales: las probabilidades de rechazar la Hipótesis Nula cuando es falsa son mayores que en las hipótesis bilaterales.

Como un resultado de $p < .05$ *suele interpretarse* como *real*, o como que merece la pena notificarlo, el establecer hipótesis de una cola o de dos colas es importante; afecta a las conclusiones del estudio (se da por probada una hipótesis con mayor facilidad) y tiene además otras consecuencias, *como el conseguir que se publique o no una investigación*. Si $p > .05$ (se acepta la Hipótesis Nula), la investigación tiene menos probabilidades de ser publicada (como nota Pillemer, 1991; es éste además un comentario que se ve en otros muchos autores).

No se trata solamente de aceptar o rechazar la Hipótesis Nula, sino que puede estar presente (al menos en algunos ambientes académicos) el deseo (o la necesidad) de publicar (*publish or perish*), o al menos está el deseo de obtener resultados significativos (todo esto es muy discutible; lo *no estadísticamente significativo* puede ser relevante y significativo con otros criterios).

Según Pillemer (1991) las buenas revistas rechazan más del 80% de los artículos y estudios que les envían para su publicación; otros autores dan estimaciones parecidas: en las mejores revistas de investigación, y más citadas, y al menos en el área de la investigación educacional, se acepta todo lo más un 10% de los trabajos enviados (Stanley, 1992).

6. Cuándo se pueden utilizar Hipótesis direccionales

1º El principio para decidir si es oportuno proponer una hipótesis direccional (unilateral) es simplemente determinar y *justificar si hay alguna razón*, alguna consideración lógica para esperar una diferencia en *una dirección*, es decir, cuando la hipótesis es *a priori* direccional; así suele indicarse en los textos y es lo que encontramos en autores bien conocidos (como McNemar, 1962, McGuigan, 1993 y muchos otros).

2º Esta *predicción de la dirección* de la diferencia hay que hacerla *antes de recoger los datos* y antes de ver los resultados. Lo que no se puede hacer es esperar a los resultados, y si no son estadísticamente significativos para una hipótesis bidireccional (o lo que es lo mismo, *no direccional*) y sí lo son para una hipótesis unidireccional, proponer *a posteriori* una hipótesis unidireccional.

Algunos autores prefieren las hipótesis unidireccionales cuando en función de los resultados se va a tomar una decisión, y los resultados son relevantes. Hoy más bien, y de cara a tomar una decisión o a dar una generalización por cierta, se tiende a *acumular resultados*, a calcular el *tamaño del efecto medio* proveniente de una serie de estudios semejantes, lo que supone ver la ciencia, al menos en este ámbito, como *cumulativa*: no hay *un único estudio definitivo*, sino una acumulación de datos que nos indican en *qué dirección* van los resultados y cuál es su *magnitud*.

7. Qué vemos en la práctica

7.1. Vemos a veces hipótesis unilaterales y otras veces (con más frecuencia posiblemente) hipótesis bilaterales, sin embargo (como nota Pillemer, 1991 y cita numerosos ejemplos) es corriente ver en artículos de investigación los valores de p ($p < .05$ por ejemplo) *sin especificar* si se trataba de una o de dos direcciones.

7.2. También es frecuente ver hipótesis bilaterales (porque se especifica así o se deduce de los datos que se ofrecen) cuando a primera vista el planteamiento parece que pide una hipótesis unilateral como decíamos en la introducción (se espera que una muestra concreta tenga una media mayor, por ejemplo cuando se comparan los grupos extremos en el análisis de ítems; o cuando se espera que el método A produzca *mejores* resultados, no simplemente resultados distintos a los del método B).

7.3. En ocasiones las hipótesis unilaterales parecen *forzadas* por el deseo de obtener resultados significativos con más facilidad (cuando realmente se puede discutir si hay base para una hipótesis de este tipo).

Por ejemplo, Lee y Bryk (1986, 1989) *prueban* que los alumnos que atienden a colegios para un solo sexo aprenden mejor que los que van a colegios mixtos. Utilizan pruebas de una cola porque su hipótesis es *a priori* unidireccional; lo que esperan es que los alumnos funcionen mejor en escuelas no mixtas. Marsh (1989a) arguye que algunas de las diferencias que favorecen a los alumnos de colegios para un solo sexo no hubieran sido significativas si se hubieran utilizado hipótesis bidireccionales y que *no hay base suficiente* para establecer hipótesis de una cola; las diferencias significativas en cualquier dirección tienen interés y pueden tener implicaciones políticas (Marsh, 1989b).

Este es uno de los problemas con las hipótesis unidireccionales: se favorecen determinadas conclusiones cuando es cuestionable en primer lugar que la hipótesis inicial se deba proponer como hipótesis unidireccional. Lo que sucede es que hay hipótesis unidireccionales que parecen muy lógicas y justificables, pero la vida va a veces por otros caminos y contradice lo que nos parecía incuestionable.

8. Problemas con las Hipótesis unidireccionales

El que las hipótesis unilaterales presentan problemas se ve en el hecho de que no hay o no ha habido acuerdo generalizado entre los investigadores; son muchas las citas que se pueden traer. Fue un tema muy discutido en los años 50 y 60 (Steger, 1971, recoge los principales artículos en los que se defienden las dos posturas) y en parte lo sigue siendo (citas de Pillemer, 1991, de finales de los años 80); los argumentos vienen a ser los mismos. En la práctica se advierte al menos una postura de *cautela* ante las hipótesis de una cola (unidireccionales) que ya no suelen utilizarse.

Resumimos brevemente los problemas que presentan las hipótesis unidireccionales.

8.1. Problemas de interpretación cuando los resultados son estadísticamente significativos pero en la dirección opuesta a la especificada

El primer problema está en que podemos obtener con más facilidad una diferencia estadísticamente significativa, pero en dirección contraria. Este resultado no previsto es, por lo menos, *incómodo*. McNemar (1962) advierte que si los resultados son estadísticamente significativos y en dirección opuesta, el investigador debe por lo menos sentirse mal (*have a red face*), porque los resultados no son consistentes ni con la hipótesis alterna ni con la hipótesis nula (no es lo mismo no demostrar que demostrar lo contrario...). Kirk (1995) comenta que las hipótesis unilaterales *castigan* al investigador que no señala bien la dirección de la diferencia. Para Dallal (2001) ningún investigador está dispuesto a aceptar como no significativo un resultado estadísticamente significativo pero en dirección contraria⁹.

Una solución es modificar la hipótesis de una cola y convertirla en otra de dos colas sobre la marcha, con lo que el resultado no será estadísticamente significativo y se evitan problemas de interpretación. Pero esta solución no es académicamente seria; la hipótesis inicial hay que mantenerla tal como se formuló. Cuando se establece una hipótesis unilateral, lo que se decide desde el principio es que un resultado en dirección contraria *no tiene mayor interés*, y se considera dentro de lo que puede suceder aleatoriamente (un resultado significativo *en dirección contraria* supone aceptar la hipótesis nula si la hipótesis era unilateral; realmente el resultado *no es* estadísticamente significativo).

Cuando con hipótesis de una cola los resultados son negativos (estadísticamente significativos y en dirección opuesta), la interpretación es que no se ha confirmado la hipótesis propia (no ha habido éxito en el intento de rechazar la hipótesis nula), y no se puede aceptar como conclusión que se ha probado la hipótesis contraria. Con estas hipótesis se introduce un *sesgo a favor* de la confirmación de las propias hipótesis, porque no se pueden confirmar las hipótesis opuestas.

Lo que puede suceder es que la lógica de la hipótesis unidireccional puede también tener sentido si el resultado es en dirección contraria. Si la hipótesis es que con el método A la media será más alta que con el método B, si resulta mejor el método B, este resultado puede ser de interés. Lo que está claro es que los resultados estadísticamente significativos y en dirección contraria a la prevista son incómodos y plantean dudas en la interpretación. Si en contra de lo especificado previamente con el método B (o en su caso la ausencia de método, de terapia, etc.) se obtienen mejores resultados, es incómodo afirmarlo simplemente porque se formuló una hipótesis unidireccional. Siendo consecuentes con este tipo de hipótesis, no se puede considerar probado o confirmado nada en sentido contrario a pesar de la aparente evidencia de los datos.

8.2. Las Hipótesis direccionales y el descubrimiento por sorpresa

Las hipótesis direccionales nos cierran el paso a posibles descubrimientos de interés.

a) Los resultados en la investigación educacional y psicológica *no son siempre coherentes*; es frecuente encontrar en la literatura experimental muchos resultados *contradicторios* (un análisis ya clásico de las inconsistencias en los resultados de estudios semejantes es Jackson, 1980). Estas contradicciones, habituales en los resultados de estudios empíricos, nos dicen que las teorías que subyacen a estas predicciones unidireccionales, y las generalizaciones que se

⁹ "This (*el poner en 1.64 el punto crítico*) makes one-sided tests very attractive to those whose definition of success is having a statistically significant result... What damns one-tailed tests in the eyes of most statisticians is the demand that *all* differences in the unexpected direction--large and small--be treated as simply non significant. I have never seen a situation where researchers were willing to do this in practice... It is surprising to see one-sided tests used in the 21st century" (Dallal, 2001, en *One-Sided Tests*).

hacen, no son lo suficientemente poderosas como para aventurarse en predicciones unidireccionales.

Lo *prudente* es formular hipótesis bilaterales o (con más propiedad) *no direcciones*. Es lo que hemos advertido antes (Marsh, 1989b): las diferencias significativas en cualquier dirección tienen siempre interés, y con frecuencia no hay base suficiente para establecer *a priori* hipótesis unidireccionales, aunque se piense que se tienen buenas razones. En este sentido *las hipótesis bilaterales son más informativas*: cualquier resultado (una diferencia significativa en cualquiera de las dos direcciones) es más aprovechable y se interpreta con más facilidad.

b) La lógica de la hipótesis de una cola es incompatible con el *aprender con hallazgos poco usuales*, inesperados... Resultados inesperados pueden llevar a nuevas teorías, a otros análisis. Estos resultados inesperados (significativos y en dirección contraria a la prevista) no son una excepción, son de gran interés y pueden llevar a nuevos conceptos o reformulaciones. Este tipo de objeciones a las hipótesis de una cola no son de ahora (por ejemplo, Burke, 1953 y Jones, 1954, los dos reproducidos en Steger, 1971).

Lane (2003) también recuerda que el mayor interés de un resultado está precisamente cuando contradice las expectativas, de manera que cuando un investigador establece hipótesis unidireccionales y se compromete a ignorar los efectos o resultados en una dirección para fijarse solamente en la otra dirección, puede verse forzado a escoger entre ignorar un resultado potencialmente importante o utilizar las técnicas de inferencia estadística de manera poco honrada¹⁰.

9. Qué dicen los diversos autores

Ya hemos advertido la polémica y la falta de acuerdo en este punto, sobre si conviene o no conviene formular hipótesis unidireccionales. De alguna manera están enfrentadas la lógica del modelo estadístico (que justifica las hipótesis de una cola) y la *praxis* (son hipótesis que crean problemas de interpretación).

Veamos lo que dicen unos pocos autores relevantes o lo que encontramos en textos muy conocidos. Estas opiniones no constituyen una *muestra aleatoria* de autores, pero reflejan la postura, a veces de cierta ambivalencia, que se mantiene en este tema. En general todos mantienen la *legitimidad teórica* de los dos tipos de hipótesis; las dos son coherentes con el modelo de inferencia estadística en el que se basan, pero notamos un *trato diferente* en la exposición de los ambos tipos de hipótesis. Cuando tratan de hipótesis unidireccionales suele recomendarse al menos *cierta prudencia*. Ordenamos cronológicamente (según la edición consultada) esta breve selección de autores.

1. Eysenck (1960, p. 271) no es ambivalente; es muy radical cuando afirma que *el cálculo de los valores de p en hipótesis de una cola no tiene lugar en psicología*

2. McNemar (1962, pp. 62-63) recomienda hipótesis de una cola si nuestra predicción es en una dirección, pero, como ya hemos mencionado, también advierte que hay que tener *cautela* (la palabra *cautela* es frecuente en este contexto); la hipótesis tiene que ser hecha *antes* de recoger los datos, y recuerda la *embarazosa situación* para el investigador si los resultados son significativos y en dirección opuesta (realmente si son en dirección opuesta, no pueden considerarse estadísticamente significativos...). Al menos indirectamente no parece recomendar las hipótesis de una cola.

¹⁰ “Frequently the most interesting aspect of an effect is that it runs counter to expectations. Therefore, an experimenter who committed him or herself to ignoring effects in one direction may be forced to choose between ignoring a potentially important finding and using the techniques of statistical inference dishonestly” (Lane, 2003).

3. Glass y Stanley (1974, la primera edición inglesa es de 1970) señalan, como hacen otros autores, que en la década de los años 50 y 60 uno de los temas de moda fue discutir sobre las ventajas e inconvenientes de las pruebas de una y de dos colas. La opinión de estos autores es que existe una *alta probabilidad* de utilizar *inadecuadamente* los contrastes unidireccionales (p. 287).

4. Guilford y Fruchter (1973, pp. 173-174) notan el *desacuerdo* entre investigadores: no todos están de acuerdo en si se deben formular hipótesis de una cola; la aceptan cuando hay *razones lógicas*.

5. Runyon y Haber (1984) mantienen la postura tradicional pero, de nuevo, *con cautela*. La hipótesis alterna puede ser *ocasionalmente* unidireccional pero tiene que ser 1) formulada *a priori*, 2) fundada en razones, y 3) antes de recoger los datos. En este caso la hipótesis nula también es unidireccional; si la Hipótesis Alterna es $P > Q$, la Hipótesis Nula será $P \leq Q$. Si rechazamos la Hipótesis Nula, podemos aceptar la validez de la Hipótesis Alterna. En definitiva aceptan la legitimidad de las hipótesis unilaterales, pero advierten que su uso debe ser más bien *occasional* (no habitual).

Estos autores insisten además en lo que ya hemos indicado previamente: que realmente nunca aceptamos la Hipótesis Nula en un sentido estricto (en el sentido de afirmar que es realmente verdadera); simplemente fracasamos en el intento de rechazarla (p. 206). La Hipótesis Nula o la rechazamos o *fracasamos* en el intento de rechazarla.

6. Morh (1992), ya mencionado, reconoce que muchas hipótesis deberían ser lógicamente *unidireccionales* (un método *es mejor que* otro, etc.) pero que en la práctica es habitual utilizar los valores de t de las pruebas bidireccionales; en estos casos el nivel usual de $\alpha = .05$ es realmente de $.025$ y es lo que debería manifestarse si hay una hipótesis clara unidireccional previa.

7. Kirk (1995:58) en su excelente texto de psicología experimental advierte que en la ausencia de suficiente información, el investigador debe ir a lo seguro (*play it safe*) y utilizar pruebas de *dos colas* (y consecuentemente hipótesis *no direcciones*).

8. Stockburger (1996), autor de un conocido texto *on line*, recomienda utilizar pruebas bilaterales en caso de duda; una razón que aduce es que los que revisan los artículos presentados para su publicación suelen *desconfiar* de las pruebas unilaterales¹¹.

9. Para Dallal (2001) no es admisible que *todas* las diferencias, grandes o pequeñas, en la dirección no esperada sean tratadas como no significativas (*It is surprising to see one-sided tests used in the 21-st century*, Dallas, 2001, en *One-Sided Tests*)

10. Lane (2003) reconoce que los tests unidireccionales no se utilizan habitualmente y que si no se dice lo contrario debe asumirse que la prueba es bidireccional.

Podemos concluir afirmando que en general la mayoría de los autores de textos importantes recomiendan por principio hipótesis bidireccionales, y cuando se explica lo que podemos denominar *doctrina tradicional* sobre hipótesis de una cola (razones fundadas y *a priori*, antes de recoger los datos), son también bastantes los autores que de una manera u otra recomiendan *cautela*. Podríamos preguntarnos qué se quiere decir con la repetida palabra *cautela* o expresiones equivalentes; podemos interpretar que *en general* no se recomienda el formular

¹¹ “Because reviewers of articles submitted for publication are sometimes suspicious when a one-tailed t-test is done, the recommendation is that if there is any doubt, a two-tailed test should be done” (Stockburger, 1996).

hipótesis de una cola aunque *aparentemente* haya razones para hacerlo así (si no las hay, y muy claras, sobra la *cautela*...).

10. Qué hacer en la práctica

10.1. Preferencia por las hipótesis no direccionales

Possiblemente una postura cautelosa (muy refrendada en la práctica de lo que se ve publicado) es la más indicada, y la cautela en este caso quiere decir que lo prudente es utilizar pruebas de dos colas como criterio general; ya hemos visto que para muchos las hipótesis de una sola cola no son admisibles.

a) *En temas discutidos es preferible una postura más bien conservadora*

Es claro que las hipótesis unilaterales, aunque se puedan justificar, presentan problemas; al menos el argumento que podemos llamar de *autoridad* va en esa línea. Es una razón aducible para atenerse a *hipótesis bilaterales* como norma habitual. Si nos equivocamos al no rechazar la hipótesis nula, lo hacemos a favor de la postura más segura.

b) *Las hipótesis direccionales se fundamentan con más dificultad*

No se trata ya de la mera autoridad de unos autores; las razones expuestas para evitar, o ser al menos cautelosos, en el uso de hipótesis unilaterales son muy válidas. Así lo demuestran los *resultados contradictorios* de muchos estudios, o al menos el hecho de que con frecuencia no se confirmen las hipótesis que otros han confirmado. Esto muestra la *dificultad en fundamentar bien hipótesis unilaterales*. Esta situación hay que verla como normal: cualquier comparación, experimento, etc., depende de muchas variables situacionales, las relaciones que se dan en una situación no se dan en otra, es imposible o muy difícil controlar todas las variables relevantes, etc.; no hay que admirarse de resultados contradictorios o no confirmatorios. Consecuentemente hay que *desmitificar la posibilidad teórica de establecer fácilmente hipótesis unilaterales muy certeras*.

c) *Las hipótesis no direccionales son más informativas*

Las hipótesis bilaterales no ofrecen especiales problemas de interpretación: un resultado no anticipado, una diferencia en la dirección opuesta a la esperada, se interpreta sin conflicto con el modelo de inferencia utilizado. Es más, la experiencia nos dice que lo inesperado no solamente es *normal*, sino que se presta a nuevas reflexiones e hipótesis. Las sorpresas y hallazgos que contradicen nuestras expectativas nos ayudan a buscar nuevas explicaciones, a matizar nuestros planteamientos, etc.

Además si tenemos una perspectiva más amplia de lo que es la investigación en las ciencias sociales, y tenemos en cuenta las posibilidades del *meta-análisis* en el que cada estudio particular puede contribuir a una futura síntesis más matizada, estos resultados inesperados, y a veces contra toda lógica aparente, son muy útiles. En su momento, por ejemplo, se puede descubrir que todos los estudios con resultados inesperados tenían algo en común que explica ese resultado atípico.

d) *Con hipótesis direccionales se obtienen con más facilidad resultados estadísticamente significativos pero también con frecuencia irrelevantes*

Resultados estadísticamente significativos y a la vez irrelevantes se pueden obtener con cualquier tipo de hipótesis, pero esto es más probable que suceda con hipótesis unilaterales porque se requieren valores más bajos de t (en el caso del contraste de medias). Esto lo vemos advertido ya desde hace tiempo (por ejemplo, Blalock, 1960; Nunnally, 1960 y Kirk, 1996, nos recuerdan que cuando no se rechaza la hipótesis nula es por falta de sujetos). Basta observar las fórmulas para caer en la cuenta de que si sube el número de sujetos aumentará el valor de t. Por

esta misma razón una variable que produce diferencias estadísticamente significativas en pequeñas muestras hay que tenerla más en cuenta (Blalock, 1960, p. 126). El peligro asociado a estos valores significativos obtenidos con facilidad en muestras grandes es la frecuencia con que se confunde la *significación estadística* con la *relevancia conceptual* o *importancia práctica*.

La confusión entre significación estadística y relevancia conceptual o práctica no es algo exclusivo de este tipo de hipótesis; simplemente se agudiza un problema de interpretación que es común en la verificación de hipótesis.

Podemos añadir que con hipótesis en las que es más fácil rechazar (o *no aceptar*) la hipótesis nula, se hacen más presentes los problemas propios de este paradigma: dar excesiva importancia a los niveles de significación y descuidar otros aspectos metodológicos, incluso cuantitativos, dar una importancia indebida al experimento único en vez de poner el énfasis en la acumulación de resultados, descuidar el cálculo y la interpretación del *tamaño del efecto* que nos cuantifica la magnitud de la diferencia independientemente de que sea o no sea estadísticamente significativa, etc.

10.2. No limitarse a calcular e interpretar, como último dato, los valores relacionados con la significación estadística (*t*, *F*, *p*)

a) Desmitificar los niveles de probabilidad

El valor de *p*, como dato final de interés para exponer y comentar conclusiones, es muy limitado. En primer lugar no hay nada sagrado acerca de la convención aceptada (y razonable si no nos quedamos ahí) de que si concluimos que $p < .05$ podemos dar por probada nuestra hipótesis (*seguramente Dios ama el .06 tanto como el 0.5*, Rosnow y Rosenthal, 1984). Poner todo en la dicotomía $p < .05$ ó $p > .05$ no tiene ninguna base, se le ha dado excesiva importancia y limita las posibilidades de interpretación (Pillemer, 1991). Nunnally (1960), y no es el único, tiene afirmaciones muy radicales sobre este tema: *el modelo de comprobación de hipótesis peor utilizado es el de la hipótesis nula, el énfasis en la hipótesis nula es poco informativo, en la vida real la hipótesis nula casi nunca es verdadera* (se refiere de manera más específica a la facilidad con que se rechaza la hipótesis nula al aumentar el tamaño de la muestra).

Sobre estos valores de *p* hay que tener en cuenta *qué dicen* (probabilidad de ocurrencia en caso de no relación o no diferencia) y *qué no dicen* (magnitud, relevancia... en una situación dada). Después de todo, estos valores de *p* no dicen mucho (o no dicen nada) sobre la *importancia* de una diferencia en una situación dada. Un valor de *p* interpretado como demasiado grande ($p > .05$, no rechazaríamos la hipótesis nula) lo que nos cuestiona es el poder generalizar este resultado a la población, y nada más, pero puede tener importancia *local* (Cohen y Hyman, 1979). Puede haber diferencias grandes y relevantes (hablando del contraste de medias) y a la vez no significativas, y al revés, podemos encontrar diferencias *estadísticamente* significativas e irrelevantes. La clásica *diferencia estadísticamente significativa* hay que tenerla más en cuenta (con *menos urgencia* de más consideraciones) con muestras pequeñas, ya que en ese caso es más difícil encontrar diferencias significativas (Blalock, 1960, p. 126).

b) Dar más información: calcular el tamaño del efecto u otras medidas de magnitud

Este cálculo nos dice más que el *sí* o el *no* con que solemos interpretar (equívocamente) los valores de *p* (*sí* afirmo la diferencia en la población, *no* afirmo la diferencia). Las diferencias (o las relaciones) pueden ser grandes y pequeñas; el *grado* o *magnitud* es importante en la interpretación independientemente de la significación estadística. Nuestras conclusiones no deben girar exclusivamente en torno al *sí* o el *no* (*sí* o *no* diferencia significativa...), sino que deben incorporar y tener en cuenta la *magnitud* de la diferencia o de la relación (*tamaño del efecto*), y se debe examinar también que variables están asociadas a esa magnitud.

La recomendación de utilizar *siempre* el tamaño del efecto (y también los *intervalos de confianza*) es constante (por ejemplo Kirk, 1996, 2001; Hubbard y Ryan, 2000, Vacha-Haase, 2001, y se podrían mencionar muchos más). El cálculo del tamaño del efecto lo exige ya la política editorial de buenas revistas (Thompson, 1996, 2002; Hubbard y Ryan, 2000; Huberty, 2002¹²) y figura en las orientaciones (*guidelines*) de la *American Psychological Association* (Wilkinson and Task Force on Statistical Inference APA Board of Scientific Affairs, 1999; American Psychological Association, 2001).

Además de los valores de p hay que especificar los valores exactos de medias, desviaciones, etc., lo mismo que los *intervalos de confianza* (entre qué límites se encuentra la media de la población), con tanta frecuencia omitidos en la exposición de los resultados. Si además de la diferencia exponemos los intervalos de confianza de las medias, la información que damos equivale en la práctica a utilizar hipótesis de dos colas (Pillemer, 1991, p. 16).

Por lo que respecta a los valores de p , lo *deseable* es dar sus *valores exactos* (en hipótesis unidireccionales o bidireccionales, pero especificando de qué hipótesis se trata), sin limitarse a p mayor o menor de .05, .01, etc. El dar valores exactos de p es más informativo y permite otros cálculos adicionales (como combinar valores de p procedentes de varios experimentos, Rosenthal, 1982, pp. 64 y ss.). La dificultad para dar los valores exactos de p es que en las tablas de los textos no vienen todos los valores posibles de t (ó F , etc.); sin embargo las probabilidades asociadas a cualquier valor de t y de grados de libertad están ya programados en programas estadísticos (de ordenadores e incluso de calculadoras y en muchos programas disponibles en Internet).

c) Desmitificar el experimento único y poner el énfasis en la acumulación de resultados

En pura teoría, con la muestra adecuada y el diseño adecuado, nos bastaría un único experimento o estudio para llegar a conclusiones definitivas. Es obvio que esto no es lo que sucede. Lo que está detrás de todo esto es un paradigma para concluir *causalidad*: las mismas circunstancias producen los mismos efectos. Pero *las mismas circunstancias casi nunca se dan*. Las *otras explicaciones* que Campbell y Stanley (1966) codificaron en 12 fuentes de invalidez, han pasado ya a 33 (Cook y Stanley, 1976). Son muchas variables para controlarlas eficazmente. Cualquier experimento, comparación, etc., depende del contexto, de variables situacionales específicas; las relaciones que se dan una situación no se dan en otra, etc. Por eso hoy, y desde hace ya tiempo, la tendencia es a alejarnos del *mito del estudio único y decisivo* y poner el énfasis en la *acumulación de resultados* (Light y Pillemer, 1984; Rosnow y Rosenthal, 1989), y en general a utilizar todas las posibilidades y sugerencias metodológicas del *meta-análisis*.

Con los énfasis nuevos que vemos en la práctica de la investigación experimental (desmitificación de los niveles de significación y del experimento único, más énfasis en cálculos adicionales que expresan magnitud y no mera probabilidad), las hipótesis de una cola pierden valor como práctica habitual.

De todas maneras podemos preguntarnos que podemos hacer cuando se dan estas tres circunstancias:

- a) Utilizamos *hipótesis bilaterales* como es habitual (y los valores de t correspondientes, $t > 1.96$ en muestras grandes y por lo tanto $p < .05$),
- b) El resultado *no es estadísticamente significativo* ($t < 1.96$) pero
- c) La hipótesis se podría haber confirmado si la hipótesis hubiera sido unidireccional (porque $t < 1.64$) y esta unidireccionalidad *tiene sentido y se puede justificar racionalmente*,

¹² Huberty (2002) menciona 19 revistas que exigen en su política editorial el cálculo del *tamaño del efecto*.

En estos casos se puede advertir que *en el caso de haber propuesto una hipótesis unidireccional se podría haber no aceptado la Hipótesis Nula de no diferencia*. Aun añadiendo esta advertencia en estos casos, y a la vista de lo que es práctica común en la que están de acuerdo tantos autores relevantes, es preferible y más seguro atenerse *en las conclusiones* a lo que se establece en las hipótesis bidireccionales. Por otra aparte, y como ya hemos advertido, en estos casos (*t menor* que 1.96 pero *mayor* que 1.64) la diferencia no suele ser cuantitativamente relevante, sobre todo en muestras grandes. También cabe proponer *como hipótesis plausible* que, manteniendo la hipótesis bilateral, muy probablemente se hubiera obtenido una diferencia estadísticamente significativa con una muestra mayor.

11. Referencias bibliográficas

1. AMERICAN PSYCHOLOGICAL ASSOCIATION (2001). *Publication manual of the American Psychological Association* (5^a ed.). Washington D.C
2. BLALOCK, HUBERT M., (1960,) *Social Statistics*, New York, McGraw-Hill
3. BURKE, C. J., (1953), A Brief Note on One-Tailed Tests, *Psychological Bulletin*, 50, 384-387.
4. CAMPBELL, D.T. and STANLEY, J.C., (1966), *Experimental and Quasi-Experimental Designs for Research*, Chicago, Rand McNally.
5. COHEN, S.A. and HYMAN, J.S., (1979), How Come So Many Hypotheses in Educational Research Are Supported? (A Modest Proposal), *Educational Researcher*, 8, 11, 12-16.
6. COOK, T.D. and CAMPBELL, D.T. (1979) *Design and Analysis of Quasi-Experiments for Field Settings*, Chicago, Rand McNally.
7. DALLAL, GERARD E. (2001) *The Little Handbook of Statistical Practice* (en *Multiple Comparisons*) <http://www.tufts.edu/~gdallal/LHSP.HTM> (consultado 5 de Octubre, 2007).
8. EYSENCK, H. J. (1960), The concept of statistical significance and the controversy about one-tailed tests, *Psychological Bulletin*, 67, 269-271.
9. GLASS, GENE V. y STANLEY, JULIAN C., (1974), *Métodos Estadísticos Aplicados a las Ciencias Sociales*, Madrid, Prentice-Hall Internacional.
10. GUILFORD, J.P. Y FRUCHTER, B., (1973), *Fundamental Statistics in Psychology and Education*, New York, McGraw-Hill.
11. HERNÁNDEZ SAMPIERI, ROBERTO; FERNÁNDEZ COLLADO, CARLOS y BAPTISTA LUCIO, PILAR (2002). *Metodología de la Investigación*, 2^a edición. México: McGraw Hill.
12. HUBBARD, RAYMOND AND RYAN, PATRICIA A., (2000). The Historical Growth of Statistical Significance Testing in Psychology-and Its Future Prospects. *Educational and Psychological Measurement*, Vol. 60 (5), 661-681
13. HUBERTY, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, 62 (2), 227-240.
14. JACKSON, GREGG B., (1980), Methods for Integrative Reviews, *Review of Educational Research*, 50, 438-460.
15. JONES, LYLE V., (1954), A Rejoinder on One-Tailed Tests, *Psychological Bulletin*, 51, 585-586
16. KIRK, ROGER E., (1995). *Experimental Design, Procedures for the Behavioral Sciences*. Pacific Grove: Brooks/Cole.
17. KIRK, ROGER E. (1996). Practical Significance: a Concept Whose Time Has Come. *Educational and Psychological Measurement*, Vol. 56 (5), 746-759.
18. LANE, DAVID M. (2003) *HyperStat Online Textbook*
<http://www.davidmlane.com/hyperstat/index.html>
19. LEE, V. E. and BRYCK, A. S., (1986), Effects of single-sex secondary schools on student achievement and attitudes, *Journal of Educational Psychology*, 78, 381-395.

20. LEE, V. E. and BRYCK, A. S., (1989), Effects of single-sex school: response to Marsh, *Journal of Educational Psychology*, 81, 647-650.
21. LIGHT, R. L. and PILLEMER, D. B., (1984), *Summing up: The science of reviewing research*, Cambridge, MA, Harvard University Press.
22. MARSH, H. R., (1989a), Effects of attending single-sex and coeducational high schools on achievement, attitudes, behaviors, and sex differences, *Journal of Educational Psychology*, 81, 70-85
23. MARSH, H. R., (1989b), Effects of attending single-sex and coeducational schools, a response to Lee and Bryck, *Journal of Educational Psychology*, 81, 651-653
24. MCGUIGAN, F. J., (1993) *Experimental Psychology, Methods of Research*, 6th edit., Englewood Cliffs, N.J., Prentice Hall.
25. MCNEMAR, QUINN, (1962), *Psychological Statistics*, 3rd edit., New York, John Wiley and Sons, 451 pp.
26. MOHR, LAWRENCE B., (1992), *Understanding Significance Testing*, Sage University Paper, Series on Quantitative Applications in the Social Sciences, Newbury Park, Sage.
27. PILLEMER, DAVID B., (1991), One-Versus Two-Tailed Hypothesis Tests in Contemporary Educational Research, *Educational Researcher*, 20, 9, 13-17.
28. ROSENTHAL, ROBERT, (1982), Combining Results of Independent Studies, *Psychological Bulletin*, 85, (1), 185-193.
29. RUNYON, RICHARD P. y HABER, AUDREY, (1984) *Estadística para las Ciencias Sociales*, México, Fondo Educativo Interamericano.
30. STANLEY, JULIAN C. (1992). A Slice of Advice, *Educational Researcher*, 21, (8), 25-26.
31. STATSOFT, Inc. (2007). *Electronic Statistics Textbook*. Tulsa, OK: StatSoft. (Electronic Version): WEB: <http://www.statsoft.com/textbook/stathome.html> (consultado 1, 05, 2009).
32. STEGER, JOSEPH A., (1971), *Readings in Statistics for the behavioral scientist*, New York, Holt, Rinehart and Winston.
33. STOCKBURGER, DAVID W. (1996). *Introductory Statistics: Concepts, Models, And Applications* <http://www.fmi.uni-sofia.bg/vesta/IntroBook/sbk00.htm> (First Published 7/15/96, Revised 8/5/97, Revised 2/19/98)
34. THOMPSON, BRUCE (1996). AERA Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms. *Educational Researcher*, Vol. 25 (2) 26-30.
35. THOMPSON, BRUCE (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 31(3), 24-31. http://www.aera.net/pubs/er/pdf/vol31_03/aera310309.pdf
36. TOOTHAKER, LARRY E. and MILLER, LISA (1995), *Introductory Statistics for the Behavioral Sciences*, 2nd edit., Pacific Grove, Brooks/Cole.
37. TROCHIM, WILLIAM M. *The Research Methods Knowledge Base*, 2nd Edition. Internet WWW page, at URL: <http://www.socialresearchmethods.net/kb/citing.htm> (version current as of October 20, 2006).
38. VACHA-HAASE, TAMMI (2001). Statistical Significance Should Not Be Considered One of Life's Guarantees: Effect Sizes Are Needed. *Educational and Psychological Measurement*, Vol. 61 (2), 219-224.
39. WILKINSON, LELAND, AND TASK FORCE ON STATISTICAL INFERENCE APA BOARD OF SCIENTIFIC AFFAIRS (1999) Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, Vol. 54, No. 8, 594–604, <http://www.apa.org/journals/amp/amp548594.html>