Psicothema 2007. Vol. 19, nº 2, pp. 343-349
www.psicothema.com
ISSN 0214 - 9915 CODEN PSOTEG

La interacción entre factores en el análisis de varianza: errores de interpretación

Antonio Pardo, Jesús Garrido, Miguel Ángel Ruiz y Rafael San Martín
Universidad Autónoma de Madrid

Aunque el concepto de interacción entre factores en el análisis de varianza tiene un significado teórico inequívoco (y así está recogido de forma generalizada en la literatura estadística), en la investigación empírica se producen frecuentemente errores de interpretación que, en muchos casos, conducen a conclusiones incorrectas. En este trabajo se revisan 150 artículos: el 12,7% no presta atención a la interacción (bien porque no se analiza, bien porque se analiza pero no se interpreta); el 79,1% interpreta la interacción recurriendo al análisis de los efectos simples; sólo el 8,2% analiza e interpreta correctamente la interacción. Quizá los investigadores en Psicología tienden a analizar e interpretar la interacción entre factores de forma incorrecta porque los programas informáticos más utilizados (el SPSS a la cabeza) no permiten llevar a cabo las comparaciones necesarias para evaluar una interacción significativa en los diseños factoriales con grupos al azar. Para contribuir a erradicar este problema se muestra cómo diseñar algunas de las comparaciones lineales que permiten aislar el efecto de la interacción y se explica cómo utilizar el SPSS para llevar a cabo esas comparaciones.

Interaction in ANOVA: Misconceptions. Although interaction in analysis of variance has an unequivocal theoretical meaning (and so it appears in the statistic literature), frequent misconceptions are found in empirical research, which, in many cases, lead to wrong conclusions. In this paper, 150 articles are reviewed: in 12.7% of them, no attention is paid to the interaction (either because it is not analysed or because it is analysed but not discussed); in 79.1%, interaction is studied through simple effects analysis; and only in 8.2% of the cases, interaction is correctly discussed. It could be that psychology researchers tend to analyse and interpret the interaction between factors incorrectly because the most widespread statistic packages (with SPSS in the lead) do not allow performing the comparisons needed to analyse a significant interaction in factorial designs with randomized groups. In order to contribute to eradicating this problem, we herein show how to design some of the linear comparisons that allow isolating the interaction effect, and we explain how to use SPSS to compute these comparisons.

Los modelos factoriales de análisis de varianza son ampliamente utilizados como estrategia de análisis de datos en muy diversas áreas de conocimiento. Entre las ventajas que justifican el uso tan extendido de estos modelos, quizá la más destacable sea que ofrecen la posibilidad de estudiar el efecto de la interacción entre factores.

El concepto de interacción está explícita y abundantemente tratado en la literatura estadística y, al menos desde el punto de vista teórico, tiene un significado inequívoco (véanse, por ejemplo, Hays, 1994, pp. 479-480; Jaccard 1998, pp. 3-6; Keppel y Wickens, 2004, pp. 198-206; Maxwell y Delaney, 2004, pp. 277-280; Winer, Brown, y Michels, 1991, pp. 296-298; etc.). Sin embargo, cuando se trata, no de *definir* la interacción, sino de *analizarla* e *interpretarla*, no parece que las cosas estén tan claras.

Fecha recepción: 31-5-06 • Fecha aceptación: 28-11-06
Correspondencia: Antonio Pardo
Facultad de Psicología
Universidad Autónoma de Madrid
28049 Madrid (Spain)
E-mail: antonio.pardo@uam.es

Según veremos: (1) en los manuales de diseño y análisis de datos y en los artículos que discuten de forma explícita el concepto de interacción no parece existir un acuerdo generalizado sobre la estrategia que debe seguirse para interpretar correctamente una interacción significativa; y (2) la mayoría de los trabajos de investigación revisados contienen errores de distinta naturaleza tanto en la forma de analizar la interacción como, sobre todo, en la de interpretarla.

Ya en 1970, Marascuilo y Levin (véase también Levin y Marascuilo, 1972) alertaron sobre lo que dieron en llamar el error de tipo IV y que definieron como la interpretación incorrecta del rechazo correcto de una hipótesis nula. Al definir este tipo de error se refirieron explícitamente a las estrategias utilizadas para interpretar una interacción significativa en el contexto del análisis de varianza. Varios años después, Rosnow y Rosenthal (1989b) llegaron a afirmar que la interacción es «probably the universally most misinterpreted empirical result in psychology» (p. 1.282). En la actualidad, la interacción entre factores sigue siendo, según tendremos ocasión de comprobar, un resultado frecuentemente mal interpretado.

Este trabajo tiene el doble objetivo de: (1) identificar el tipo de errores que se cometen cuando se analiza e interpreta el efecto de la interacción entre factores; y (2) proponer una estrategia para ayudar a los investigadores a analizar e interpretar correctamente ese efecto.

El concepto de interacción

El concepto de interacción entre factores admite varias formulaciones, todas ellas equivalentes. Desde un punto de vista *no formal*, decimos que existe interacción entre dos factores cuando el efecto de uno de ellos sobre la variable dependiente no es el mismo en todos los niveles del otro factor (véanse, por ejemplo, Everitt y Howell, 2005, pp. 930-931; Kirk, 1995, p. 367; o Maxwell y Delaney, 2004, p. 278). Esto equivale a afirmar que existe interacción cuando el resultado de la combinación de dos factores difiere de la suma de los efectos principales de esos factores (véanse, por ejemplo, Everitt y Howell, 2005, p. 931; Maxwell y Delaney, 2004, pp. 279-280; o Winer, Brown, y Michels, 1991, p. 296).

Para poder presentar una definición *formal* de la interacción, consideremos la notación propuesta en la tabla 1 para la configuración de un diseño factorial 2×3, con los factores A y B (con niveles A_j y B_k; j= 1, 2; k= 1, 2, 3), la media total (μ), las medias marginales (μ_{j+} y μ_{+k}) y las medias de las casillas (μ_{jk}).

En el modelo de ANOVA de dos factores de efectos fijos, el efecto de la interacción se define como (véase, por ejemplo, Winer, Brown, y Michels, 1991, p. 318):

$$(\alpha\beta)_{jk} = \mu_{jk} - \mu_{j+} - \mu_{+k} + \mu \qquad [1]$$

De acuerdo con esta definición, existe interacción cuando $(\alpha\beta)_{jk} \neq 0$ para algún j o k; y no existe interacción cuando $(\alpha\beta)_{jk} = 0$ para todo j y k. Ahora bien, existen al menos dos maneras alternativas de interpretar la definición [1] (véase Jaccard, 1998, pp. 3-10):

1. Como la *desviación que experimentan las medias de las casillas respecto de los efectos principales de los factores*:
   1.a. No interacción: $\mu_{jk} = \mu_{j+} + \mu_{+k} - \mu$ (para todo j y k).
   1.b. Interacción: $\mu_{jk} \neq \mu_{j+} + \mu_{+k} - \mu$ (para algún j o k) [2].
   Según esta definición, existe interacción cuando la media de una o más casillas no es función directa de sus respectivas medias marginales. Esto significa que, si existe interacción, el efecto de *la combinación* de los factores A y B difiere de *la suma* de los efectos de A y B.
2. Como *diferencias entre las medias de las casillas y las medias marginales*:
   2.a. No interacción: $\mu_{jk} - \mu_{j'k} = \mu_{j+} - \mu_{j'+}$ (para todo j, j' o k).
   2.b. Interacción: $\mu_{jk} - \mu_{j'k} \neq \mu_{j+} - \mu_{j'+}$ (para algún j, j' o k) [3].
   Según esta definición, existe interacción cuando la diferencia entre las medias de dos casillas de la misma columna (lo mismo vale para las filas) no es igual que la diferencia entre sus correspondientes medias marginales.

Tabla 1
Notación utilizada en un diseño 2×3

	B_1	B_2	B_3	
A_1	μ_{11}	μ_{12}	μ_{13}	μ_{1+}
A_2	μ_{21}	μ_{22}	μ_{23}	μ_{2+}
	μ_{+1}	μ_{+2}	μ_{+3}	μ

Las expresiones [2] y [3] se deducen, ambas, de [1]; por tanto, son equivalentes. En efecto, según la expresión [2], cuando no existe interacción:

$$\mu_{11} = \mu_{1+} + \mu_{+1} - \mu \quad \text{y} \quad \mu_{21} = \mu_{2+} + \mu_{+1} - \mu$$

De donde se sigue:

$$\mu_{11} - \mu_{1+} - \mu_{+1} = -\mu \quad \text{y} \quad \mu_{21} - \mu_{2+} - \mu_{+1} = -\mu$$

Por tanto:

$$\mu_{11} - \mu_{1+} - \mu_{+1} = \mu_{21} - \mu_{2+} - \mu_{+1}$$

Lo cual lleva a:

$$\mu_{11} - \mu_{21} = \mu_{1+} - \mu_{2+}$$

que no es otra cosa que la expresión [3]. De este argumento se desprende que es irrelevante tomar [2] o [3] como referente para la discusión. Sin embargo, las interpretaciones basadas en [3] suelen resultar más fáciles de entender.

La primera formulación (basada en [2]) ha recibido especial atención por parte de Rosnow y Rosenthal (1989a, 1989b, 1991, 1995, 1996), quienes han llevado el argumento al extremo insistiendo en que para poder interpretar correctamente el efecto de la interacción hay que despojarlo de todos los elementos que incluye. Es así como se llega a las *medias residualizadas* o *residuos de interacción* que, siempre según Rosnow y Rosenthal, son los únicos que informan cabalmente sobre el efecto de la interacción. Pero no parece que este enfoque haya merecido la aceptación de todos (véase Meyer, 1991; Petty, Fabrigar, Wegener, y Priester, 1996); y tampoco parece que los investigadores estén dispuestos a incorporar a sus hábitos el plus de comprensión que exige.

Unas sencillas transformaciones permiten comprobar que la definición [3] implica que, si existe interacción, la diferencia entre los niveles A_1 y A_2 no es la misma en los tres niveles de B; y lo mismo vale decir de la diferencia entre los niveles B_1, B_2 y B_3 en los dos niveles de A. Éste es el significado de la interacción entre los factores A y B. Y ésta es la idea que se pretende destacar aquí: la interpretación de la interacción requiere *comparar diferencias*.

Cuando la situación se simplifica a un diseño 2×2, unas sencillas transformaciones permiten comprobar que la ecuación [3] equivale a:

$$\mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} \qquad [4]$$

La comparación propuesta en [4] es la que corresponde al único grado de libertad asociado a la interacción en un diseño 2×2. Por tanto, si el estadístico F asociado al efecto de la interacción es significativo, una interpretación basada en las diferencias comparadas en [4] agota el significado de la interacción, lo cual implica que no es necesario recurrir a comparaciones adicionales para interpretar una interacción significativa (debe tenerse en cuenta que, si se verifica [4], también se verifican: $\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$ y $\mu_{11} + \mu_{22} = \mu_{12} + \mu_{21}$; por tanto, una interacción significativa puede interpretarse recurriendo a cualquiera de estas tres comparaciones de diferencias).

Errores cometidos al interpretar la interacción

Para cubrir el primer objetivo de este trabajo, se han revisado los 848 artículos publicados en *Psicothema* (699 artículos entre 2000 y 2004) y en el *International Journal of Clinical and Health Psychology* (149 artículos entre 2001 y 2005). Se han elegido estas dos revistas por ser las revistas españolas de Psicología con mayor impacto acumulado en los últimos años. *Psicothema* es una de las dos revistas españolas de Psicología indexada en los últimos años en el Journal Citation Reports; y el *International Journal of Clinical and Health Psychology* es la revista de Psicología con mayor índice de impacto en los últimos años en la lista IN-RECS de revistas españolas de ciencias sociales (http://ec3.ugr.es/in-recs/).

En una primera revisión se han seleccionado los artículos en cuyos resúmenes se indicaba que se había utilizado un ANOVA de dos o más factores y aquellos que en virtud del diseño informado requerían ese abordaje analítico. Puesto que los resúmenes no siempre facilitaban información suficiente para determinar si se había utilizado o no un ANOVA factorial o si el diseño exigía su utilización, se procedió a una revisión visual del cuerpo de los artículos dudosos en busca de tablas o gráficos que sugiriesen la utilización de un diseño factorial.

En esta primera revisión de resúmenes, tablas y gráficos se seleccionaron 166 artículos (19,6% del total). Todos los artículos elegidos fueron impresos y leídos íntegramente para confirmar la utilización de un ANOVA factorial o la presencia de un diseño susceptible de ser analizado mediante la aplicación de un ANOVA factorial. Posteriormente, 16 de ellos fueron descartados por no contener ningún efecto de interacción significativo o por utilizar herramientas de análisis no paramétricas.

De los 150 artículos finalmente considerados, 15 (10,0%) pertenecen al *International Journal of Clinical and Health Psychology* y 135 (90,0%) a *Psicothema*. En esta cifra están incluidos todos los artículos que recogen alguna interacción significativa (141 artículos) más los que, habiendo utilizado un diseño factorial, no analizan la interacción y, por tanto, no es posible saber si ésta es o no significativa (9 artículos). En el resto de los artículos revisados, o no se aplica un ANOVA factorial o, si se aplica, ninguna de las interacciones alcanza a ser significativa.

Dependiendo de la forma de analizar e interpretar la interacción, los 150 artículos finalmente considerados se han clasificado en una de las siguientes cuatro categorías:

A) *No se analiza la interacción*, a pesar de que los objetivos explícitos del estudio y/o el diseño experimental lo requieren. Se incluyen aquí los artículos en los que únicamente se ofrecen resultados descriptivos (al margen del análisis de los efectos principales) y, consiguientemente, la interacción se interpreta a partir de un gráfico o de una tabla de medias (ocasionalmente se hacen afirmaciones en la discusión que sugieren la presencia de una interacción que no ha sido contrastada).

B) *Se analiza la interacción pero no se interpreta* (o se interpreta, incorrectamente, como un efecto principal). Es decir, existe una interacción significativa a la que no se le presta atención.

C) *La interacción se analiza e interpreta recurriendo a los efectos simples*. Aquí se han incluido los artículos en los que, tras una interacción significativa, se recurre al análisis de los efectos simples por separado para interpretarla, y aquellos en los que se recurre directamente al análisis de los efectos simples por separado sin valorar previamente la presencia de una interacción significativa.

D) *La interacción se analiza e interpreta correctamente*: ya sea comparando diferencias, ya sea argumentando que la combinación de efectos difiere de la suma de efectos.

Aunque en la revisión se han incluido los trabajos en los que se utiliza un diseño factorial (dos factores o más), únicamente se ha prestado atención a las interacciones de primer orden (interacciones entre dos factores). La tabla 2 ofrece el resultado de la clasificación. Puesto que algunos artículos incluyen más de un experimento, la suma de las clasificaciones llevadas a cabo es mayor (158) que el número de artículos revisados (150). En los artículos que contienen más de una interacción significativa, si todas ellas se interpretan de la misma manera se ha emitido un solo juicio; si hay distintas formas de interpretarla, se han emitido tantos juicios como formas de interpretarla. En los apartados que siguen se explica cada categoría de clasificación y las razones por las cuales las categorías A, B y C se consideran estrategias incorrectas o erróneas.

No se analiza o no se interpreta la interacción (categorías A y B)

En los trabajos clasificados en la categoría A no se utiliza el estadístico F ni ninguna otra estrategia analítica para determinar si el efecto de la interacción es estadísticamente significativo (representan el 5,7% del total de clasificaciones). De todos los trabajos en los que se da esta omisión cabe distinguir entre los que simplemente no prestan atención a la interacción, es decir, los que ni la analizan ni la interpretan, y los que, aun no analizándola, la interpretan a partir del gráfico de líneas o del tamaño de las medias de las casillas.

En la categoría B se han clasificado los trabajos que, habiendo analizado la interacción y habiendo resultado significativa, no la interpretan (7,0% del total).

Con la única excepción de las escasas ocasiones en las que las hipótesis de un estudio se refieren a comparaciones planeadas, no prestar atención a la interacción en un diseño factorial (lo que se hace en los trabajos clasificados en las categorías A y B) constituye un desacierto evidente: dado que una interacción significativa está indicando que el efecto de un factor no es el mismo en todos los niveles del otro, puede afirmarse que el significado de los efectos principales de un diseño queda matizado (incluso alterado) por la presencia de una interacción significativa. En este sentido, aunque algunos autores sugieren que, siendo significativa la interacción, todavía podría tener sentido interpretar los efectos principales en determinadas circunstancias, no por ello dejan de señalar que la interacción debe ser interpretada (véase, por ejemplo, Howell, 2002, p. 432; Keppel y Wickens, 2004, p. 244; León y Mon-

Tabla 2
Clasificación de las interacciones en función del tratamiento que reciben

Categoría (tratamiento que recibe la interacción)	***n***	**%**
A) No se analiza	9	5,7
B) Se analiza pero no se interpreta	11	7,0
C) Se analiza e interpreta recurriendo al análisis de los efectos simples	125	79,1
D) Se analiza e interpreta correctamente	13	8,2

tero, 2001, p. 165). Y la postura más generalizada recomienda no prestar atención a los efectos principales en presencia de una interacción significativa (Games, 1973, p. 305; Kirk, 1995, p. 370; Maxwell y Delaney, 2004, p. 301; Pedhazur y Pedhazur, 1991, p. 523; Winer, Brown, y Michels, 1991, pp. 326-327).

Por tanto, no prestar atención a la interacción en un diseño factorial (bien porque no se analiza, bien porque no se interpreta) supone cometer un error consistente en desechar información que no sólo es relevante en sí misma, sino que, además, podría estar modificando sensiblemente el significado de los efectos principales.

Por otro lado, interpretar la interacción a partir de las medias muestrales (o de la representación gráfica de esas medias) sin valorar su significación estadística supone olvidar un aspecto crucial del análisis de datos: los resultados muestrales podrían estar reflejando, no diferencias reales, sino simplemente fluctuaciones atribuibles al azar muestral (es decir, se podría estar considerando significativo un efecto que podría no serlo).

La interacción se interpreta analizando los efectos simples (categoría C)

La categoría C es la que más se repite. En los trabajos clasificados en esta categoría, la interacción se interpreta recurriendo al análisis de cada efecto simple por separado. Este tipo de error se comete en el 79,1% de los artículos revisados. Veamos en qué nos basamos para afirmar que esta estrategia es errónea.

Consideremos un diseño 2×2. En la ecuación [4] se ha definido la interacción como una diferencia de diferencias. Ahora bien, la diferencia $\mu_{11} - \mu_{21}$ es el *efecto principal simple* (o *efecto simple*) del factor A en el nivel B_1; y la diferencia $\mu_{12} - \mu_{22}$ es el *efecto simple* de A en B_2. Por tanto, *afirmar que existe efecto de la interacción equivale a afirmar que el efecto simple de A en B_1 difiere del efecto simple de A en B_2*. Pero justamente a partir de aquí es donde surge el problema, porque la afirmación referida a la diferencia entre los dos efectos simples se interpreta con frecuencia de esta incorrecta manera: *si al analizar los dos efectos simples del factor A por separado se comprueba que uno de ellos es significativo y el otro no, se puede concluir que los efectos simples del factor A no son iguales en los dos niveles del factor B* (también se utiliza este argumento en sentido inverso: si los efectos simples no son iguales, entonces el efecto de la interacción es significativo).

¿Por qué es incorrecta esta interpretación? Porque se está afirmando que son distintas dos cosas que no se han comparado; es decir, se está afirmando que el efecto simple de A en B_1 difiere del efecto simple de A en B_2 sin haber comparado entre sí ambos efectos simples. Y lo cierto es que uno de los dos efectos simples de A podría ser significativo y el otro no tanto si existe interacción significativa como si no; y ambos efectos simples podrían ser significativos o no significativos tanto si existe interacción significativa como si no (véase Keppel y Wickens, 2004, p. 254).

La razón de esta aparente inconsistencia radica en el hecho de que un efecto simple incluye parte del efecto principal y parte del efecto de la interacción. En concreto, cada efecto simple de A en B_1 viene dado por (véase Kirk, 1995, pp. 377-378):

$$\alpha_j \text{ en } B_1 = \alpha_j + (\alpha\beta)_{j1} \qquad [5]$$

donde el término $\alpha_j = \mu_{j+} - \mu$ se refiere a cada uno de los J efectos asociados a los niveles del factor A y el término $(\alpha\beta)_{j1}$, ya definido en la ecuación [1], se refiere a los J términos de interacción asociados al nivel 1 del factor B. De la ecuación [5] se deduce que un efecto simple puede ser significativo bien porque es significativo el efecto principal que incluye (es decir, $\alpha_j \neq 0$ para algún j), bien porque es significativo el efecto de la interacción (es decir, $(\alpha\beta)_{j1} \neq 0$ para algún j), bien por ambas cosas.

Las implicaciones prácticas de esta afirmación son importantes. Imaginemos que el factor A define dos grupos de tratamiento (G_E = experimental y G_C = control) y que el factor B representa dos momentos en el tiempo (pre- y postratamiento). En un diseño de estas características el investigador suele estar interesado en averiguar si el tratamiento tiene algún efecto sobre el grupo experimental (por supuesto, se asume que se trata de algún efecto distinto del no-tratamiento sobre el grupo control). Para obtener esta información no basta con analizar el efecto principal del factor A, sino que es necesario comparar lo que ocurre en el postratamiento (efecto simple de A en B_2) con lo que ocurre en el pretratamiento (efecto simple de A en B_1). Ahora bien, si para realizar esta comparación se recurre al análisis de los efectos simples por separado (estrategia utilizada en los trabajos clasificados en la categoría C), puede ocurrir que, *siendo significativo el efecto de la interacción*, no haya diferencias significativas entre G_E y G_C ni en el pre- ni en el postratamiento (véase Figura 1-a); y también puede ocurrir que haya diferencias significativas tanto en el pre- como en el postratamiento (véase Figura 1-b). En la estrategia basada en el análisis de los efectos simples por separado, cualquiera de estos dos resultados llevaría a concluir que no es posible afirmar que exista efecto del tratamiento. Sin embargo, en clara discrepancia con esta conclusión, la presencia de una interacción significativa estaría indicando que la diferencia entre G_E y G_C no es la misma en el pre- y en el postratamiento; lo cual debería llevar a concluir que existe efecto del tratamiento (pues, en un diseño de estas características, una interacción significativa implica efecto del tratamiento).

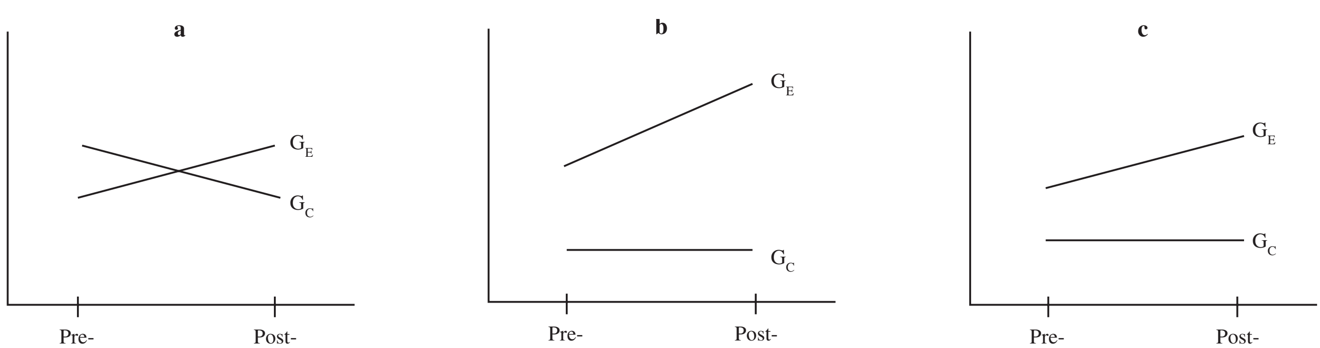

Figura 1. *Distintas pautas de interacción en un diseño 2×2*

También puede ocurrir que, *no siendo significativo el efecto de la interacción* (es decir, no habiendo diferencias entre lo que ocurre en el pre- y en el postratamiento), la diferencia entre G_E y G_C en el pretratamiento (efecto simple de A en B_1) no sea significativa y sí lo sea la diferencia entre G_E y G_C en el postratamiento (efecto simple de A en B_2). Este resultado podría llevar a afirmar que existe efecto del tratamiento cuando el hecho de que la interacción sea no significativa estaría descartando la presencia de un efecto del tratamiento (véase Figura 1-c).

Por tanto, para poder afirmar que existe efecto del tratamiento no basta con saber que G_E y G_C no difieren en el pre- y sí en el postratamiento, como tampoco basta con saber que G_E cambia entre el pre- y el postratamiento mientras que G_C no lo hace (de todo esto es de lo que informan los efectos simples y donde con mayor frecuencia se centra el análisis de los investigadores). Para poder afirmar que existe efecto del tratamiento, la diferencia observada en el post- hay que referirla a la observada en el pretratamiento (o, equivalentemente, el cambio observado en G_E entre el pre- y el postratamiento hay que referirlo al cambio observado en G_C), y esto sólo es posible hacerlo comparando diferencias, que es justamente lo que se hace cuando se analiza el efecto de la interacción.

Aunque una interacción significativa coincidirá, en muchos casos, con la presencia de efectos simples diferenciados (es decir, unos significativos y otros no), esto no tiene por qué ser necesariamente así. Por tanto, si bien el análisis de los efectos simples podría llevar a las mismas conclusiones que la interpretación correcta de la interacción, esa estrategia debe ser considerada errónea porque puede conducir a conclusiones erróneas.

No analizar la interacción, o no interpretarla, o interpretarla sin aplicar una prueba de significación (categorías A y B), son errores de difícil justificación, no ya sólo de quienes los cometen, sino de los revisores que los pasan por alto. Sin embargo, recurrir a los efectos simples para interpretar una interacción significativa es un tipo de error que, podría decirse, no es tan evidente (pues implica una confusión que se ajusta a cierta lógica) ni tan grave (pues, muchas veces, analizar e interpretar los efectos simples por separado lleva, según se ha señalado, a la misma conclusión que analizar e interpretar correctamente la interacción). No obstante, aun no siendo un tipo de error tan evidente ni tan grave como los demás, ya se ha argumentado que utilizar esta estrategia debe ser calificada de errónea porque puede llevar a conclusiones erróneas.

Y, sin embargo, se trata de una estrategia que se utiliza con altísima frecuencia (79,1% de los artículos revisados). En nuestra opinión, existen algunas razones que explican lo muy extendido que está el uso de esta estrategia. En primer lugar, a pesar del hecho ya señalado de que los *efectos simples incluyen tanto efectos principales como de interacción* está suficientemente documentado en los manuales de diseño y análisis estadístico (Kirk, 1995, pp. 377-378; Winer, Brown, y Michels, 1991, pp. 326-332), hasta el punto de que autores de la talla de Kirk han llegado a afirmar que «contrastar hipótesis sobre los efectos simples… puede ser interesante, pero no ayuda a comprender la interacción entre dos variables» (1995, p. 383), a pesar de esto, decimos, no pocos prestigiosos manuales de diseño y análisis presentan los efectos simples como la estrategia apropiada (y en algunos casos única) para interpretar los datos en presencia de una interacción significativa (Howell, 2002, pp. 432, 489; Jaccard, 1998, p. 20; Keppel y Wickens, 2004, p. 247; Maxwell y Delaney, 2004, p. 308; Pedhazur y Pedhazur, 1991, p. 509; etc.). Algo parecido sucede también con las referencias en castellano (véase, por ejemplo, León y Montero, 2001, p. 165; o Pascual, 1998, p. 97).

En segundo lugar, el programa informático que se utiliza en la inmensa mayoría de los artículos revisados (el SPSS) no permite realizar un análisis *post hoc* de la interacción (al menos, no de forma directa): permite, mediante sintaxis, analizar los efectos simples y las comparaciones por pares asociadas a los efectos simples (véase Pardo y Ruiz, 2005, pp. 374-375, 380-385); también permite comparar diferencias relacionadas con la interacción en los diseños de medidas repetidas (véase Field, 2005, pp. 471-478; esta estrategia no ha sido utilizada en ninguno de los artículos revisados); pero no permite efectuar los contrastes que aíslan el efecto de la interacción en los diseños factoriales con grupos al azar.

Para cubrir el segundo objetivo de este trabajo, en el siguiente apartado se explica cómo definir comparaciones lineales entre medias para analizar correctamente una interacción significativa en un diseño factorial con grupos al azar y cómo utilizar el SPSS (programa estadístico más utilizado por los investigadores que publican sus trabajos en las revistas de Psicología) para llevar a cabo esas comparaciones.

Cómo efectuar comparaciones para analizar la interacción

De la misma manera que es posible definir comparaciones lineales de un grado de libertad para interpretar, descomponiéndolo, un efecto principal (véase, por ejemplo, Pardo y San Martín, 1998, pp. 287-292), también es posible definir comparaciones lineales de un grado de libertad para conseguir interpretar una interacción significativa.

Y aunque el número de estas comparaciones es muy elevado (Abelson, 1997, p. 318), las comparaciones que más ayudan a los investigadores a interpretar una interacción significativa suelen ser aquellas que permiten comparar entre sí los efectos simples. Por ejemplo, en un diseño factorial 2×3 como el propuesto en la tabla 1, la necesidad de interpretar una interacción significativa quedará satisfecha, normalmente, comparando entre sí cada efecto de A en cada nivel de B, es decir, comparando entre sí los efectos simples de A (o, si se prefiere, comparando entre sí los efectos simples de B, lo cual es equivalente desde el punto de vista de las conclusiones que se alcanzan).

Ahora bien, para comparar entre sí los efectos simples de A no basta con valorar si un efecto simple es significativo y otro no para, de esta forma, decidir que son distintos (esto es, básicamente, lo que se hace en los trabajos clasificados en la categoría C; y ya se ha explicado que esta estrategia no permite aislar el efecto de la interacción porque en los contrastes de los efectos simples se está valorando, además del efecto de la interacción, parte del correspondiente efecto principal; véase ecuación [5]). Comparar entre sí los efectos simples de A requiere:

1. Comparar la diferencia entre μ_{12} y μ_{22} (o efecto simple de A en B_2) con la diferencia entre μ_{11} y μ_{21} (o efecto simple de A en B_1);
2. Comparar la diferencia entre μ_{13} y μ_{23} (o efecto simple de A en B_3) con la diferencia entre μ_{11} y μ_{21} (o efecto simple de A en B_1);
3. Comparar la diferencia entre μ_{13} y μ_{23} (o efecto simple de A en B_3), con la diferencia entre μ_{12} y μ_{22} (o efecto simple de A en B_2).

Es decir, comparar entre sí los efectos simples de A requiere efectuar estas tres comparaciones:

$\psi_1 = (\mu_{12}–\mu_{22}) – (\mu_{11}–\mu_{21})$
$\psi_2 = (\mu_{13}–\mu_{23}) – (\mu_{11}–\mu_{21})$
$\psi_3 = (\mu_{13}–\mu_{23}) – (\mu_{12}–\mu_{22})$

Ordenando se obtiene:

$\psi_1 = –\mu_{11} + \mu_{12} + \mu_{21} – \mu_{22}$
$\psi_2 = – \mu_{11} + \mu_{13} + \mu_{21} – \mu_{23}$
$\psi_3 = – \mu_{12} + \mu_{13} + \mu_{22} – \mu_{23}$

Y asignando coeficientes:

$\psi_1 = (–1)\mu_{11} + (1)\mu_{12} + (0)\mu_{13} + (1)\mu_{21} + (–1)\mu_{22} + (0)\mu_{23}$
$\psi_2 = (–1)\mu_{11} + (0)\mu_{12} + (1)\mu_{13} + (1)\mu_{21} + (0)\mu_{22} + (–1)\mu_{23}$
$\psi_3 = (0)\mu_{11} + (–1)\mu_{12} + (1)\mu_{13} + (0)\mu_{21} + (1)\mu_{22} + (–1)\mu_{23}$

Por supuesto, estas tres comparaciones son sólo algunas de las que es posible definir para intentar interpretar una interacción significativa, pero tienen más utilidad de la que aparentan. Por ejemplo, no es infrecuente que un investigador esté interesado en analizar una tendencia concreta. Una situación típica de este tipo es aquella en la que la diferencia entre los niveles del factor *A* va aumentando (o disminuyendo) conforme progresan los niveles del factor *B* (resultado muy típico de los ensayos clínicos en los que se aplica un tratamiento y se hace seguimiento), es decir:

$$\psi_4 = (\mu_{11}–\mu_{21}) < (\mu_{12}–\mu_{22}) < (\mu_{13}–\mu_{23})$$

Esta comparación representa la interacción entre el factor *A* y el componente lineal del factor *B*, es decir, representa una situación en la que la diferencia entre los niveles del factor *A* se va incrementando linealmente a través de los niveles del factor *B* (por supuesto, podría trabajarse con un incremento cuadrático utilizando la misma lógica). Los coeficientes que permiten definir esta comparación (véase, por ejemplo, Kirk, 1995, pp. 392-394) se obtienen multiplicando (interacción) los coeficientes que definen la diferencia entre los niveles de *A* (coeficientes de ψ_A) por los que definen una tendencia lineal en los niveles de *B* (coeficientes de $\psi_{B(\text{lineal})}$). La tabla 3 recoge los coeficientes correspondientes a ψ_A y $\psi_{B(\text{lineal})}$ (estos últimos se pueden encontrar en cualquier tabla de polinomios ortogonales), y el producto entre estas dos columnas arroja los coeficientes que recoge la columna $\psi_{A \text{ x } B(\text{lineal})}$. Ahora bien, estos coeficientes se corresponden con los ya utilizados previamente en la comparación ψ_2 para comparar los efectos simples de *A* en B_1 y B_3. Es decir, la comparación ψ_4 es equivalente a la ψ_2.

Una característica interesante de las comparaciones ψ_1, ψ_2 y ψ_3 es que los coeficientes que las definen permiten comprobar que todas ellas son independientes de los efectos principales (dos comparaciones son independientes si la suma de los productos de sus coeficientes vale cero; véase, por ejemplo, Kirk, 1995, pp. 115-116). Los coeficientes de la tabla 3 permiten comprobar, por ejemplo, que al sumar los productos de los coeficientes de ψ_1 y ψ_A se obtiene:

$$(–1)(1)+(1)(1)+(0)(1)+(1)(–1)+(–1)(–1)+(0)(–1) = 0$$

Con los coeficientes de las comparaciones ψ_2 y ψ_3 se obtiene el mismo resultado. Y también se obtiene el mismo resultado si se toma como referente el efecto principal de *B*. Sin embargo, esto no ocurre con las comparaciones que definen los efectos simples (ya se ha señalado que un efecto simple incluye tanto parte de la interacción como del efecto principal; véase ecuación 5). Por ejemplo, la columna $\psi_{A|B1}$ de la tabla 3 recoge los coeficientes correspondientes a la comparación que define el efecto simple de *A* en B_1. Sumando los productos de estos coeficientes por los que definen el efecto principal de *A* se obtiene:

$$(1)(1)+(0)(1)+(0)(1)+(–1)(–1)+(0)(–1)+(0)(–1) = 2$$

Este resultado confirma, por otra vía, la afirmación de la ecuación [5] de que los efectos simples no son independientes de los efectos principales.

Para contrastar hipótesis del tipo $\psi_h = 0$ puede utilizarse el estadístico *t* convencional (véase Kirk, 1995, pp. 137-143):

$$t = \frac{\hat{\psi}_h}{\sqrt{MCE \sum_k \frac{c_k^2}{n_k}}} \qquad [6]$$

aplicando la correspondiente corrección para controlar la tasa de error asociada al número de comparaciones que se lleven a cabo. Este estadístico se distribuye según el modelo de probabilidad *t* de *Student* con grados de libertad igual al número de casos menos el número de casillas (*h* se refiere a una comparación cualquiera; *k* se refiere al número de casillas del diseño; *MCE* es la media cuadrática error o error cuadrático medio).

Lo que interesa destacar de esta forma de abordar el estudio de la interacción entre dos factores es que el estadístico propuesto en [6] está disponible en el SPSS dentro del procedimiento *ANOVA de un factor*. Ahora bien, para poder utilizar este procedimiento con un diseño de dos factores es necesario crear una nueva variable cuyos valores sean el resultado de combinar los valores de los factores *A* y *B*. Por ejemplo, si el factor *A* tiene dos niveles (con códigos 1 y 2) y el factor *B* tiene 3 (con códigos 1, 2 y 3), la nueva variable deberá tener 6 códigos distintos: 1 para la combinación A_1B_1, 2 para A_1B_2, 3 para A_1B_3, 4 para A_2B_1, 5 para A_2B_2 y 6 para A_2B_3. Una vez creada la nueva variable, basta con introducirla como factor en el mencionado procedimiento y definir en el cuadro de diálogo *Contrastes* las comparaciones que se desea llevar a cabo (véase Pardo y Ruiz, 2005, pp. 354-359, para una explicación de cómo llevar a cabo este tipo de comparaciones).

Por supuesto, esta estrategia es posible utilizarla porque la media cuadrática error (*MCE*) que se obtiene con el modelo factorial original es exactamente la misma que se obtiene cuando se realiza la transformación para aplicar el modelo de un factor (puesto que la *MCE* representa la variabilidad existente dentro de las casillas, su valor no cambia porque se altere el orden de las mismas).

Tabla 3
Coeficientes para comparaciones lineales en un diseño 2×3

Casillas	ψ_1	ψ_2	ψ_3	ψ_A	$\psi_{B(\text{lineal})}$	$\psi_{A \text{ x } B(\text{lineal})}$	$\psi_{A\|B1}$
A_1 B_1	–1	–1	0	1	–1	–1	1
A_1 B_2	1	0	–1	1	0	0	0
A_1 B_3	0	1	1	1	1	1	0
A_2 B_1	1	1	0	–1	–1	1	–1
A_2 B_2	–1	0	1	–1	0	0	0
A_2 B_3	0	–1	–1	–1	1	–1	0

Discusión y conclusiones

1. A pesar de que el concepto de interacción entre factores en el contexto del análisis de varianza factorial tiene un significado teórico inequívoco (y así está recogido de forma generalizada en la literatura estadística), en la investigación empírica se producen frecuentemente errores de interpretación que, en muchos casos, conducen a conclusiones incorrectas.
2. De los 158 juicios emitidos, el 12,7% se refiere a artículos en los que no se presta atención a la interacción (bien porque no se analiza, bien porque se analiza pero no se interpreta). Y los juicios más frecuentes (79,1% del total) se refieren a artículos en los que la interacción se interpreta recurriendo al análisis de los efectos simples (ya se ha señalado que esta estrategia debe ser considerada errónea porque puede conducir —aunque no necesariamente— a conclusiones erróneas).
3. Quizá los investigadores en Psicología tienden a analizar e interpretar la interacción entre factores de forma incorrecta porque los programas informáticos más utilizados (el SPSS a la cabeza) no permiten llevar a cabo las comparaciones necesarias para evaluar una interacción significativa en los diseños factoriales con grupos al azar. Para contribuir a erradicar este problema, se ha mostrado cómo diseñar algunas de las comparaciones lineales que permiten aislar el efecto de la interacción (en concreto, las comparaciones que habitualmente interesa realizar a los investigadores en la mayor parte de los diseños que aplican) y se ha explicado cómo utilizar el SPSS para llevar a cabo esas comparaciones.

Referencias

Abelson, R.P., y Prentice, D.A. (1997). Contrast tests of interaction hypotheses. *Psychological Bulletin*, *2*, 315-328.

Everitt, B.S., y Howell, D.C. (2005). *Encyclopedia of statistics in behavioral science* (vol. 2, pp. 929-933). Chichester, Sussex: Wiley.

Field, A. (2005). *Discovering statistics using SPSS* (2ª ed.). Thousand Oaks, CA: Sage.

Games, P.A. (1973). Type IV errors revised. *Psychological Bulletin*, *80*, 304-307.

Hays, W.L. (1994). *Statistics* (5ª ed.). Belmont, CA: Wadsworth.

Howell, D.C. (2002). *Statistical methods for psychology* (5ª ed.). Boston: PWS-Kent Publishing Company.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Thousand Oaks, CA: Sage.

Keppel, G., y Wickens, Th. D. (2004). *Design and analysis: A researcher's handbook* (4ª ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kirk, R.E. (1995). *Experimental design. Procedures for the behavioral sciences* (3ª ed.). Belmont, CA: Brooks/Cole.

León, O.G., y Montero, I. (2001). Cómo explicar el concepto de interacción sin estadística: análisis gráfico de todos los casos posibles en un diseño 2×2. *Psicothema*, *13*, 159-165.

Levin, J.R., y Marascuilo, L.A. (1972). Type IV errors and interactions. *Psychological Bulletin*, *78*, 368-374.

Levin, J.R., y Marascuilo, L.A. (1973). Type IV errors and Games. *Psychological Bulletin*, *80*, 308-309.

Marascuilo, L.A., y Levin, J.R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of type IV errors. *American Educational Research Journal*, *7*, 397-421

Maxwell, S.E., y Delaney, H.D. (2004). *Designing experiments and analyzing data. A model comparison perspective* (2ª ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Meyer, D.L. (1991). Misinterpretation of interactions effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin*, *110*, 571-573.

Pardo, A., y Ruiz, M.A. (2005). *Análisis de datos con SPSS 13 Base*. Madrid: McGraw-Hill.

Pardo, A., y San Martín, R. (1998). *Análisis de datos en Psicología II* (2ª ed.). Madrid: Pirámide.

Pascual, J. (1998). Diseño entre grupos. En M.T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (eds.): *Métodos de investigación en Psicología*, Madrid: Síntesis (pp. 73-112).

Pedhazur, E.J., y Pedhazur, L. (1991). *Mesurement, design and analysis. An integrated approach*. Hillsdale, NJ: Lawrence Erlbuam Associates.

Petty, R.E., Fabrigar, L.R., Wegener, D.T., y Priester, J.R. (1996). Understanding data when interactions are present or hypothesized. *Psychological Science*, *7*, 247-252.

Rosnow, R.L., y Rosenthal, R. (1989a). Definition and interpretation of interaction effects. *Psychological Bulletin*, *105*, 143-146.

Rosnow, R.L., y Rosenthal, R. (1989b). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.

Rosnow, R.L., y Rosenthal, R. (1991). If you're looking at the cell means, you're not looking *only* at the interaction (unless all main effects are zero). *Psychological Bulletin*, *110*, 574-576.

Rosnow, R.L., y Rosenthal, R. (1995). «Some things you learn aren't so»: Cohen's paradox, Asch's paradigm and the interpretation of interaction. *Psychological Science*, *6*, 3-9.

Rosnow, R.L., y Rosenthal, R. (1996). Contrast and interactions redux: Five easy pieces. *Psychological Science*, *7*, 253-257.

Winer, B.J., Brown, D.R., y Michels, K.M. (1991). *Statistical principles in experimental design* (3ª ed.). New York: McGraw-Hill.