

POTENCIA ESTADÍSTICA DEL DISEÑO DE SOLOMON

José Fernando García Pérez, Dolores Frías Navarro y Juan Pascual Llobell
Universitat de València

El artículo describe la problemática asociada al análisis estadístico del Diseño de Cuatro Grupos de Solomon. Campbell y Stanley (1963) afirmaron que “el diseño de cuatro grupos de Solomon tiene un alto prestigio y representa el primer acercamiento serio al problema de la validez externa”. Sin embargo, la aplicación del diseño de cuatro grupos de Solomon ha sido escasa, quizá a causa de la falta de certeza acerca de su tratamiento estadístico (Braver y Braver, 1988). Estos últimos autores han añadido a las anteriores una aproximación meta-analítica que no está exenta de algunas limitaciones, especialmente las referidas al control del error tipo I.

Statistical power of Solomon design. This article describes the use and statistical treatment of the Solomon Four Experimental Design. Campbell and Stanley (1963) noted “the Solomon Four Experimental Design has higher prestige and represent the first explicit consideration of external validity factors”. Broad applicability of the Solomon Four Experimental Design has been limited in research (Braver y Braver, 1988). Perhaps the most important reason why the Solomon design is underused is the lack of certainty concerning the proper statistical treatment of this rather complicated design. Braver & Braver (1988) proposed a meta-analysis approach to test the efficacy of an intervention in the Solomon Four Design. In their article Braver and Braver provide a rationale for the sequence of decisions at each point in the algorithm, but make no attempt to estimate the effect of sequential tests on the experiment-wise error rates.

La aplicación del Diseño de Cuatro Grupos de Solomon (Solomon, 1949) es especialmente aconsejable en aquellas investigaciones donde resulta imprescindible comprobar la posible sensibilización previa de los sujetos al tratamiento (véase un ejemplo en Sprangers y Hoogstraten, 1989). El diseño, en su estructura lógico/formal (véase tabla 1), consta de cuatro grupos, de los que

dos reciben el tratamiento y los otros dos tienen la consideración de grupos de control y de todos ellos, la mitad pasan por una situación de pretest y la otra mitad no. Huck y Sandler, (1973, p. 54) han señalado que la exposición al pretest incrementa (o disminuye) la sensibilidad de los sujetos al tratamiento, aumentando (o disminuyendo) los efectos estimados de éste, lo que resulta decisivo a la hora de poder generalizar los resultados. Este inconveniente, que afecta tanto al diseño pretest/posttest con grupo de control como al diseño con grupo de control y sólo posttest, es superado en el diseño de cuatro grupos de Solomon al añadir dos

Correspondencia: José Fernando García Pérez
Facultad de Psicología
Universitat de València
46010 Valencia (Spain)
E-mail: fernando.garcia@uv.es

nuevos grupos que al no pasar por la condición de pretest permiten contrastar el posible efecto de sensibilización previa al tratamiento. Mediante este recurso metodológico, el diseño no solo gana en validez interna sino en validez externa, por lo que algunos autores llegaron a afirmar en su momento que es el “*más destacable de todos los diseños experimentales*” (Helmastadter, 1970, p-110).

Tabla 1
Estructura formal del Diseño de Cuatro Grupos de Solomon

Grupo	Pretest	Tratamiento	Posttest
1 ⇒ R	O ₁	X	O ₂
2 ⇒ R	O ₃	-	O ₄
3 ⇒ R		X	O ₅
4 ⇒ R		-	O ₆

donde:

- R ≡ aleatorización
- O ≡ observación
- X ≡ grupo de tratamiento
- ≡ grupo sin tratamiento (grupo control)

Sin embargo y a pesar de estas ventajas, el diseño no ha sido ampliamente utilizado quizá porque nunca fue suficientemente estudiado desde el punto de vista estadístico, según afirman Braver y Braver, (1988). Efectivamente, el diseño permite varias alternativas de análisis complementarios o alternativos entre sí.

En principio, admite ser analizado mediante la aplicación del Análisis de Covarianza (ANCOVA), mediante un diseño de medidas repetidas (diseño mixto), a través del Análisis de varianza (ANOVA) sobre el posttest, o mediante un diseño de bloques siempre que sea posible categorizar la variable pretest y, finalmente, mediante el análisis de las puntuaciones de ganancia/diferencia. De todas las anteriores soluciones,

probablemente la menos aconsejable sea la última de ellas. Thorndike y Hagen (1977) argumentaron contra esta última opción al comprobar que a medida que la correlación entre pretest y posttest se va acercando al valor de fiabilidad del test, la fiabilidad de las puntuaciones de diferencia se acerca a 0, lo que se observa en la tabla 2.

Tabla 2
Fiabilidad de dos pruebas

Correlación Pre-Post	0.60	0.80	0.90
0.60	0.00	0.50	0.75
0.80		0.00	0.50
0.90			0.00

Dejando, pues, de lado esta última opción, ¿hay alguna razón para elegir entre las alternativas restantes? ¿La elección se ha de contemplar como un proceso secuencial entre las distintas alternativas? ¿Cabe añadir alguna nueva?

Siendo consecuentes con las posibilidades del diseño, la primera opción lógica y necesaria a tener en cuenta sería la del diseño factorial mixto, en virtud del cual se nos posibilita descubrir la existencia del efecto de interacción, es decir del efecto de sensibilización de los sujetos al tratamiento, si realmente existe. Sólo después de descartar dicho efecto tendría sentido proceder a practicar dos estimaciones independientes del efecto experimental: vía ANCOVA sobre puntuaciones pretest/posttest y vía ANOVA sobre las puntuaciones del posttest. A estas alternativas se debería de añadir la solución meta-analítica propuesta por Braver y Braver, en 1988, sobre todo cuando ninguna de las anteriores soluciones debido a la falta de potencia estadística ha logrado detectar resultados significativos.

La aportación de Braver y Braver (1988) al análisis del diseño se fundamenta sobre el siguiente razonamiento. Puesto que el diseño permite dos estimaciones independientes

del mismo efecto experimental (vía ANCOVA y vía ANOVA), por añadidura posibilita también la aplicación de la técnica meta-analítica. Aunque ésta sea un técnica que inicialmente se ha aplicado a la integración de resultados procedentes de estudios distintos, no hay razones para presumir que no sea igualmente aplicable a estimaciones independientes procedentes de un mismo trabajo de investigación. Efectivamente, los autores ilustraron mediante un ejemplo ficticio que el resultado meta-analítico obtenido a partir de la estimación z de Stouffer (Stouffer, Suchman, DeVinney, Star y Williams, 1949) sobre los resultados de dos test estadísticos (ANCOVA y ANOVA) era estadísticamente significativo, $p = 0.040$. Concluían, en consecuencia, con la afirmación que la técnica del meta-análisis tiene mayor potencia ($z_{meta} = 1.65 + 1.25/\sqrt{2} = 2.05$, $p = 0.040$) que los análisis de covarianza ($p = .099$) y varianza ($p = .212$), considerados por separado.

Nosotros hemos aplicado el mismo razonamiento a un cuerpo de datos reales llegando a las mismas conclusiones que Braver y Braver (1988) aunque en este caso el procedimiento meta-analítico utilizado fue el propuesto por Edgington (1972) que estima la probabilidad combinada de varios resultados a partir de los diferentes niveles de significación asociados a los estadísticos originales aplicando la fórmula siguiente:

$$\bar{p} = \frac{\left(\sum_i^n p_i\right)^n}{n!}$$

¿Solución meta-analítica o artefacto estadístico?

¿Hasta qué punto la solución meta-analítica propuesta por Braver y Braver es adecuada? Hace algún tiempo aplicamos las sugerencias de estos autores a los resultados empíricos obtenidos en investigaciones tendientes a comprobar el efecto de los estados

emocionales sobre la retención inmediata (Pascual, Estarelles y Camarasa 1988). La tarea propuesta a los sujetos, tras inducir emocionalmente a los sujetos experimentales mediante la técnica de Velten (1968), fue análoga a la propuesta por Daneman y Carpenter (1983) para medir la memoria en funcionamiento. Los resultados mostraron que aún siendo eficaz la inducción emocional los rendimientos de los sujetos no presentaban diferencias estadísticamente significativas (véase Tabla 3).

Medias de rendimiento por cada categoría		
Alegre	Neutral	Depresivo
30.88	35.25	32.38

Aplicado el ANCOVA, el análisis no fue estadísticamente significativo ($p = 0.20$) y obtenida la razón t (o F) sobre las puntuaciones posttest tampoco se alcanzó el nivel convencional de significación ($p = 0.13$). Sin embargo, aplicando la solución meta-analítica propuesta por Edgington (1972), anteriormente descrita, resultó que nuestros datos obtenían una probabilidad de 0.05, por tanto estadísticamente significativa.

Alcanzada la barrera mítica del 0.05 de significación estadística es obligado preguntarse si el resultado no es más que un artefacto estadístico. Poco sabemos acerca de la verdadera distribución del estadístico de Edgington, pero lo que resulta llamativo es que el promedio de dos probabilidades bastante altas alcanza significación estadística.

En la tabla 4 hemos simulado un conjunto de probabilidades independientes (desde el nivel de significación 0.05 hasta el 0.50, tanto en filas como en columnas) y se puede comprobar que la solución meta-analítica propuesta por Edgington aplicada a la integración de dos probabilidades independientes de 0.15 genera resultados estadísticamente significativos. Cuando menos los

resultados que hemos obtenido son sorprendentes y la estimación en todo caso cabe considerarla de excesivamente liberal. ¿Es más potente la solución meta-analítica o es excesivamente liberal, sesgada a favor del rechazo de la hipótesis de nulidad? Los datos de la tabla 4 parecen sugerir esta segunda opción.

Solución meta-Analítica y tasa de error

Según la propuesta original de Braver y Braver (1988) (véase Figura 1) el test de covarianza y el tests de varianza se tienen que aplicar sólo después de haber demostrado la no existencia de interacción entre tratamiento y pretest.

Al aplicar la secuencia condicional propuesta por estos autores se debe asumir un aumento considerable de la probabilidad de cometer error *Tipo I*. Efectivamente, el error cometido en el caso de la prueba *t* sería de 0.13, para el caso de la covarianza sería de 0.11 y, finalmente, para la solución meta-analítica, la última siempre de la cadena posible, sería de 0.13. Se corre, en consecuencia, el riesgo de casi triplicar el nivel *alpha* asociado al conjunto de contrastes realizados (*tasa de error por familia de comparaciones*) al considerar que cada una de las so-

luciones analíticas está condicionada a la anterior y dependiendo, lógicamente, la tasa de error del lugar que se ocupe en la cadena de decisión (1).

Por tanto, se puede afirmar que, aunque la estrategia meta-analítica aparece a primera vista como más potente, no por ello está exenta de ciertos problemas. Concretamente hemos hablado de dos: uno, que los procedimientos analíticos mediante los que generamos la integración de resultados, concretamente el propuesto por Edgington (1972) pueden estar sesgados positivamente; y un segundo, que la solución meta-analítica exige, so pena de hacer de la misma una estimación excesivamente liberal, controlar la tasa de error tipo I asociada a la misma. Ninguna de las dos consideraciones formales aquí descritas se ha tenido en cuenta hasta el presente, lo que exige un estudio más profundo del diseño de cuatro grupos de Solomon desde los criterios de potencia de la prueba y control del error tipo I.

Nota

Modo de obtención del error Tipo I, según formulación de Bonferroni = $1 - (1 - \alpha)^c$ donde *c* es el numero de contrastes efectuados y α el error Tipo I que se asume en cada comparación individual.

Tabla 4
Distribución \bar{p} con dos muestras

<i>Pr</i>	0.050	0.060	0.070	0.080	0.090	0.100	0.150	0.200	0.250	0.300	0.400	0.500
0.050	0.005	0.006	0.007	0.003	0.010	0.011	0.020	0.031	0.045	0.061	0.101	0.151
0.060	0.006	0.007	0.008	0.010	0.011	0.013	0.022	0.034	0.048	0.065	0.106	0.157
0.070	0.007	0.008	0.010	0.011	0.013	0.014	0.024	0.036	0.051	0.068	0.110	0.162
0.080	0.008	0.010	0.011	0.013	0.014	0.016	0.026	0.039	0.054	0.072	0.115	0.168
0.090	0.010	0.011	0.013	0.014	0.016	0.018	0.029	0.042	0.058	0.076	0.120	0.174
0.100	0.011	0.013	0.014	0.016	0.018	0.020	0.031	0.045	0.061	0.080	0.125	0.180
0.150	0.020	0.022	0.024	0.026	0.029	0.031	0.045	0.061	0.080	0.101	0.151	0.211
0.200	0.031	0.034	0.036	0.039	0.042	0.045	0.061	0.080	0.101	0.125	0.180	0.245
0.250	0.045	0.048	0.051	0.054	0.058	0.061	0.080	0.101	0.125	0.151	0.211	0.281
0.300	0.061	0.065	0.068	0.072	0.076	0.080	0.101	0.125	0.151	0.180	0.245	0.320
0.400	0.101	0.106	0.110	0.115	0.120	0.125	0.151	0.180	0.211	0.245	0.320	0.405
0.500	0.151	0.157	0.162	0.168	0.174	0.180	0.211	0.245	0.281	0.320	0.405	0.500

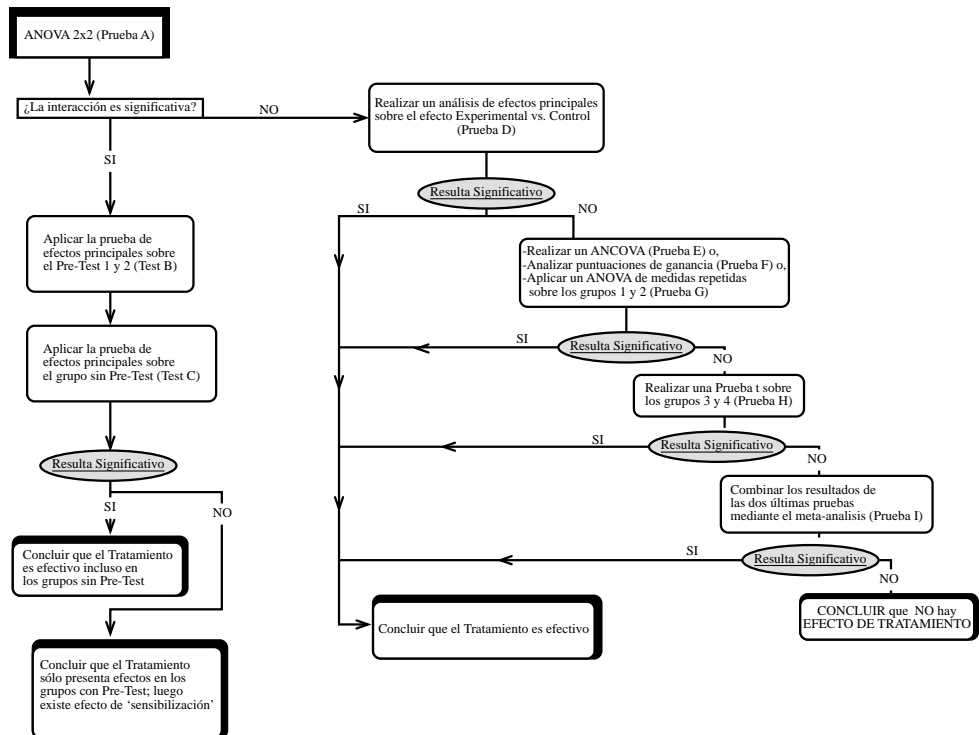


Figura 1. Adaptado a partir de Braver y Braver (1988), *Statistical treatment of the Solomon four-group design: A meta-analytic approach. Psychological Bulletin*, 104(1), 150-154.

Referencias

Braver, M. C., y Braver, S. L. (1988). Statistical treatment of the Solomon Four-Groups Design: A meta-analytic Approach. *Psychological Bulletin*, 104, 150-154.

Campbell, D. T., y Stanley, J. C. (1963). *Experimental and quasiexperimental design for research*. Chicago: Rand McNally. (Traducido al castellano en 1973).

Daneman, M., y Carpenter, P. A. (1983) Individual differences in integrating information between and within sentema. *Journal of Experimental Psychology*, 9, 561-84.

Edgington, R. E. (1972). An additive method for combining probability values from independent experiments. *Journal of Psychology*, 80, 351-363.

Helmastadter, G. C. (1970). *Research concepts in human behavior: Education, psychology, sociology*. NY: Appleton-Century-Croft.

Huck, S., y Sandler, H. M. (1973). A note on the Solomon 4-groups design: Appropriate statistical analysis. *Journal of Experimental Education*, 42, 54-55.

Pascual, J., Estarells, R., y Camarasa, C. (1993). Statistical Analysis of the Solomon Four-Group design: A case study. En R. Steyer, K. F. Wender y K. F. Widaman (Eds.). *Psychometric Methodology* (pp. 402-407). Stuttgart and New York: Gustav Fisher Verlag.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, 46, 137-150.

Sprangers, M. y Hoogstraten, J. (1989). Pre-testing effects in retrospective pretest-posttest designs. *Journal of Applied Psychology*, 74, 265-272.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. y Williams, R. M., Jr. (1949). *The American soldier: adjustment during Army life* (Vol. 1). Princeton, NJ: Princeton University Press.

Thorndike, y Hagen (1977). *Measurement and evaluation in psychology and education*. New York: Wiley.

Velten, E. (1968). A laboratory task for induction of mood states. behavior. *Research and Therapy*, 6, 473-482.

Aceptado el 11 de septiembre de 1998