anales de psicología, 1991, 7 (1), 85-104

Estimacion de la precision de un diseño multivariable de medidas repetidas

Jaime Arnau Gras(*)
Angel Blanco Villaseñor
José Luis Losada López

Universidad de Barcelona

Resumen: Los diseños de investigación de medidas repetidas pueden ampliarse a estructuras más complejas si introducimos un factor de clasificación de los sujetos a fin de controlar su posible efecto y conseguir una mayor precisión en la estimación de los tratamientos. La teoría de la generalizabilidad asume que hay otras fuentes de variación además de las diferencias individuales y permite integrar cada una de estas fuentes de variación en una estructura global que admite aplicaciones particulares de la teoría estadística del muestreo.

Palabras clave: Diseños de medidas repetidas; teoría de la generalizabilidad.

Title: Accuracy stimation of a multivariable repeated measures design.
Abstract: In order to control their possible effect and to obtain greater reliability on the estimation of treatments, the repeated measures design could be extended to more complex structures when introduces a classification factgor of the subjects. Generizability theory assumes that other sources of variance in addition of the individual differences, permits to integrate each one of tghese sources of variance in a general structure that admit specific application of the sample statistical theory.

Key words: Generizability theory; repeated measures design.

Diseño de investigación de medidas repetidas: Introducción

Las técnicas de análisis de la variancia (AVAR) han constituido los procedimientos clásicos para la inferencia de hipótesis en contextos experimentales. Como sabemos estas técnicas fueron desarrolladas inicialmente por Fisher (1925), al poner las bases para su aplicación con el desarrollo de la teoría del muestreo. No obstante las técnicas analíticas elaboradas por Fisher se aplicaron básicamente a estudios de una sola variable criterio (variable dependiente), de modo que cabe señalar que las técnicas basadas en la comparación de las variancias se aplican usualmente en experimentos univariables.

El matemático estadístico Wilks (1932) destacó que los principios del análisis del AVAR podían ser fácilmente generalizables a situaciones multivariables. Es decir situaciones donde se registran simultáneamente dos o más indicadores de la variable dependiente. Aunque los

(*) **Dirección**: Facultad de Psicología. Departamento de Metodología de las Ciencias del Comportamiento. Universidad de Barcelona. 08028 Barcelona (España).

verdaderos intentos de llevar a cabo dicha generalización los encontramos, de hecho, en los trabajos de Bartlett (1947) y Tukey (1949).

A pesar de que existe una clara analogía entre el Análisis univariable de la variancia y el Análisis multivariable de la variancia (AMVAR) como ha destacado Bock (1975), la aplicación de este último análisis posee claras ventajas, dado que puede inferirse el efecto de los tratamientos sobre el conjunto de variables dependientes o indicadores considerados o tomados simultáneamente. Desde este punto de vista el AMVAR permite tener en cuenta las posibles interrelaciones entre las distintas variables dependientes. De ahí que sea un procedimiento mucho más apropiado para el estudio de los fenómenos psicológicos o variables dado que suelen estar interconexionadas dentro de una amplia red de conductas o patrones conductuales.

Si bien tradicionalmente la estructura multivariable es aquella que, como se ha dicho, tiene en cuenta una multiplicidad de variables criterio (variables dependientes) de modo que para cada sujeto se toman más de una observación o registro, no obstante tenemos una disposición de carácter experimental que si bien no se ajusta de una manera exacta a este esquema puede ser considerado como multivariable. Nos referimos a los diseños de medidas repetidas, que tienen como estructura básica el hecho de que se registra una misma conducta o indicador conductual a lo largo de una serie de puntos de observación o intervención. Estrictamente hablando esta disposición experimental no podría ser considerada multivariable, dado que es la misma conducta que se mide a lo largo de los distintos niveles de la variable dependiente. Debe tenerse en cuenta que la disposición multivariable requiere la medida de diferentes indicadores conductuales. No obstante desde un punto de vista formal el diseño de medidas repetidas puede ser considerado como una estructura multivariable y ser sometido a los análisis multivariables típicos. Dicha disposición experimental suele resolverse utilizando la técnica clásica de la variancia (AVAR) aunque para ello se requiera el cumplimiento de unas suposiciones altamente restrictivas (homogeneidad y simetría).

Una clara alternativa de análisis al AVAR clásico es sin duda el análisis multivariable de la variancia que como se ha indicado incorpora en su estructura la posible relación entre las variables y por tanto puede ser considerado como una técnica más adaptable a la situación multivariable.

Los diseños de medidas repetidas pueden ampliarse en estructuras más complejas cuando se introduce un factor de clasificación de los sujetos, a fin de controlar su posible efecto y conseguir una mayor **precisión en la estimación** de los tratamientos. De esta forma tenemos los diseños conocidos por Split-Plot que, si bien inicialmente se aplicaron a la investigación agrícola, parecen ser claramente apropiados dentro del ámbito psicológico y clínico.

El diseño Split-Plot suele utilizarse en experimentos de laboratorio para el estudio de dos factores diferentes. Como afirma Dayton (1970), los esquemas de investigación que incorporan esta estrategia son equivalentes a los diseños parcialmente anidados, siendo las referencias obligadas para un conocimiento exhaustivo de su estructura los trabajos de Cochran y Cox (1957) y Federer (1955).

Ilustración mediante un ejemplo

Si, por ejemplo, un investigador desea estudiar el efecto de dos factores de tratamiento A y B a partir de una determinada muestra debería, en primer lugar, asignar los sujetos, de acuerdo con un criterio completamente aleatorio, a los distintos niveles de la variable A (o grupos experimentales). En segundo lugar, aplicar también al azar los niveles de la variable B a los diferentes subgrupos formados dentro de A. Los niveles de la variable A son conocidos por "tratamientos de plots completos" (whole plot treatments) y, los niveles de la variable B por "tratamientos subplot" (subplot treatments). Atendiendo, por tanto, a esta estructura de diseño se tiene una primera variable cuyos valores son asignados al azar a los bloques (en este caso sujetos) y, una segunda variable cuyos valores o niveles son asignados, también al azar, dentro de cada bloque. Como destaca Fleiss (1986), las comparaciones que implican el factor de plots completos o factor de bloques suelen ser, por lo general, menos precisas y potentes que las comparaciones que implican el factor de subplots (tratamientos intra-bloques) o la interacción entre los dos factores.

La aplicación de esta estrategia de investigación dentro del contexto de experimentación clínica, consiste en caracterizar como factor de bloques (whole plot factor) a una variable de clasificación en lugar de una variable de tratamiento y, administrar todos los niveles o tratamientos del factor de subplots a cada uno de los sujetos del experimento (en cuyo caso, puede ser considerado como un factor intra-sujetos de medidas repetidas). Esta estructura experimental suele recibir, dentro del contexto metodológico, el nombre de Diseño Split-plot de medidas repetidas.

Supongamos que nos interesa estudiar el desarrollo escolar de un grupo de niños, a lo largo de una serie de cursos escolares. Para ello se elige una muestra de niños (clasificados según el sexo: cuatro niños y cuatro niñas) y se toman puntuaciones de vocabulario a través de una prueba escalada (Arnau, 1990). Las puntuaciones escaladas de vocabulario se registran al término de los siguientes años escolares: 3º, 4º, 5º y 6º de E.G.B. Los niños han sido categorizados en función del sexo.

Tabla 1: Matriz de datos del diseño.

	SUJETOS	CURSO:	3	4	5	6
V	1		1	2	3	5
A	2		2	3	4	5
R	3		2	2	3	5
Ó	4		3	4	4	5
N						
H	5		2	4	4	5
E	6		2	3	3	3
M	7		1	2	4	4
B	8		2	3	4	4
R						
A						

De los resultados obtenidos, observamos que la interacción (plots-subplots; es decir, variable entre-sujetos por variable intra-sujetos) no es, en este ejemplo, significativa ($P < .078$). Por esta razón, interpretaremos por separado los efectos de estos dos componentes o factores.

Si analizamos, en primer lugar, el efecto entre-grupos (plots completos), que en nuestro caso es la variable sexo, se infiere que no es una variable predictora de las puntuaciones que obtienen los sujetos en vocabulario. Por otra parte, el efecto debido a la variable curso (sub-plots) resulta significativa ($P < .000$). Por lo tanto existe una diferencia significativa entre los distintos cursos y, en consecuencia, se infiere la efectividad de dicha variable sobre los sujetos.

Tabla 2: Cuadro resumen del análisis de la variancia.

FUENTES DE VARIACION	SUMA DE CUADRADOS	GL	CUADRADOS MEDIOS	F	P
CONSTANT	32.513	1	32.513	26.907	0
PLOTS	12	1	12	10	759
WITHIN CELLS	725	6	121		
SUB-PLOTS	3.262	3	1.080	4.121	0

Nuestro deseo, a partir de estos supuestos de un diseño multivariable de medidas repetidas y de estos resultados, es determinar la precisión que conseguiríamos en la generalización de estos resultados a otras situaciones similares en las que dispusiéramos de un número mayor de muestras para cada uno de los efectos principales y de interacción de las diversas fuentes de variación.

Para ello trataremos de determinar las diferentes disposiciones de medida que nos brinda la teoría de la generalizabilidad y ofrecer así un marco más amplio, con la detección de los errores de medida, que facilite la interpretación de los resultados encontrados.

Ambos métodos tienen como común denominador al análisis de la variancia, aunque ambos difieren en la interpretación de los resultados que corresponden a la estimación. Lo que en principio pudiera pensarse que son semejanzas y disimilitudes (Arnau, Blanco y Losada, 1990) no es otra cosa que la realidad ofrecida en la mayoría de los trabajos de investigación que utilizan una u otra técnica (una, 'medidas repetidas', utilizada mayoritariamente en estudios experimentales y la otra, 'generalizabilidad', en estudios observacionales). Lo que nosotros pretendemos demostrar es que ambas, dependiendo del diseño de investigación, son complementarias.

Dado que la teoría de la generalizabilidad no es lo suficientemente conocida en contextos de investigación experimentales, trataremos ahora de hacer una breve exposición que nos lleve a reconocer las semejanzas y disimilitudes entre una y otra técnica. De esta forma, y en los párrafos que siguen trataremos de determinar por qué esta solución es más adecuada que la psicométrica al diseño que presentamos en nuestro ejemplo inicial.

Al final de nuestro recorrido, y una vez que se hayan examinado los diferentes planes de medida que nos ofrece la teoría de la generalizabilidad en la estimación de la precisión de cada uno de ellos, llevaremos a cabo uno de los procesos más deseables en esta precisión: Las decisiones a tomar en caso de detección de errores de medida o lo que es lo mismo la optimización de nuestro diseño para futuras investigaciones. En definitiva, las conclusiones de este trabajo nos permitirán asegurar la precisión de los resultados a otras muestras diferentes o por el contrario nos servirán para rediseñar nuestro estudio con el fin de obtener mejores éxitos en investigaciones futuras.

La decisión final corresponde a la optimización de resultados y a la maximización de los diferentes coeficientes de generalizabilidad. Para ello, y basándonos en los trabajos de Woodward & Joe (1973) y Marcoulides & Goldstein (1990), trataremos de establecer las relaciones costo-beneficio en función de restricciones (presupuestarias y de otro tipo) del diseño de investigación, con el fin de maximizar los coeficientes obtenidos y replanificarlo (si así fuera necesario), ya que nos permitirá determinar el número ideal de observaciones a realizar en el diseño inicial de medidas repetidas.

Objetivos de estimación de la precisión

La realización de este trabajo parte de la idea de considerar a los diseños de investigación como una estructura global en la planificación de investigaciones psicológicas. Ello nos ha llevado a considerar que la utilización de las técnicas estadísticas no son parcela exclusiva de determinadas actuaciones en función de los datos que presenta una investigación determinada. Más bien al contrario, pensamos en la complementariedad de las mismas, como es el caso al que nos referimos.

Basándonos en las recientes incorporaciones que han llevado a cabo los diseñadores del paquete de programas BMDP, en su última revisión PC90 (Dixon, Brown, Engelman, & Jennrich, 1990) tratamos de estimular la investigación psicológica a través de su propuesta en el programa BMDP 8V, que nos presenta la estimación de los componentes de variancia en los diseños de investigación de medidas repetidas. Estos hechos y la creciente utilización del diseño en la evaluación nos hace pensar en la necesidad que tienen investigadores de otras áreas por la optimización de los recursos de investigación así como en la maximización de sus resultados en función de la relación costo-beneficio. De esta forma, la teoría de la generalizabilidad permite aportar soluciones a estas necesidades y concretar la variabilidad que se produce a medida que vamos incorporando facetas (o variables, o factores) a nuestro diseño de investigación. Así determinamos la importancia de cada una de ellas en relación con las demás a través de los componentes de variancia.

El diseño de investigación de medidas repetidas nos aporta el plan de acción a considerar para la consecución de nuestros objetivos. La estimación de la precisión a través de la teoría de la generalizabilidad conlleva conocer el efecto de inclusión de nuevas facetas en el diseño, así como la posibilidad de determinar si la generalización de los resultados es precisa en el caso de disponer de un número mayor de muestras en cada uno de los efectos principales y de interacción. El programa BMDP 8V nos ayuda a interconexionar ambas cosas, a través de aplicaciones particulares de la teoría estadística del muestreo. La aplicación de todo ello a la evaluación es uno de los objetivos que pretendemos desde estas páginas.

Teoría de la generalizabilidad: Introducción

Este ejemplo de aplicación de la teoría de la generalizabilidad viene a ilustrar la ampliación que con respecto a un diseño multivariable de medidas repetidas, implica el principio de "simetría de las facetas (o variables)" de un Plan de Observación (Cardinet, Tourneur & Allal, 1976).

Suele ser difícil en ciencias del comportamiento encontrar una unidad física para medir la ejecución de una actividad (exceptuando estudios de percepción o de tiempos de reacción). Los sistemas de medida se basan, fundamentalmente, en la clasificación de los sujetos respecto a un comportamiento particular. A través del análisis factorial se tiene una garantía de que estos sistemas de medida, son representativos del conjunto de comportamientos que se pretende investigar. Se pone a punto una prueba, se estandariza y se eligen aquellas cuestiones que diferencian mejor la ejecución de las actividades por parte de los sujetos.

En el ejemplo que proponemos, los individuos (clasificados según el sexo) aprenden a ejecutar la actividad en diferentes cursos académicos (medidas repetidas en el diseño multivariable). Ahora bien, dicha solución psicométrica encierra ciertos inconvenientes (Cardinet, 1987): En el contexto escolar no parece correcto recurrir a la clasificación para construir una escala de medida, ya que el cálculo tendría que ser muy exacto. Por ello es preferible utilizar medidas absolutas y no relativas, dado que su interpretación es más directa y más clara. Tampoco tiene mucho interés que la situación de observación esté estandarizada, puesto que el conocimiento del número exacto de problemas de una prueba que es capaz de resolver un individuo en un tiempo fijo y en unas condiciones particulares no suele aportar mucha información. Más bien interesaría estimar aproximativamente su probabilidad de contestar a otros problemas presentados de forma diferente. No se trata, pues, de conocer las preguntas más difíciles que diferencian mejor a los individuos, ya que en tal caso mediríamos los límites del aprendizaje y no las adquisiciones reales. Por último, no sería un objeto de estudio crear una jerarquía de valor entre los alumnos que serviría para fundamentar una jerarquía social en detrimento de la función educativa de la escuela.

Esta es una visión particular desde la teoría de la generalizabilidad que ha sido avalada especialmente por Cardinet (1987) y que ha supuesto importantes consideraciones metodológicas en el estudio de la observación directa del comportamiento a través de la metodología observacional (Blanco, 1989). Estamos, por tanto, de acuerdo con Cardinet (1987) en que este tipo de evaluación no debe suponer una competición entre alumnos, sino una forma de seguir individualmente la progresión de cada uno de ellos y revelar sus progresos. Ello supone la comparación de niveles de ejecución sucesivos en el curso del aprendizaje (medidas repetidas en el diseño multivariable que ilustramos) y la distinción entre progreso real y fluctuaciones aleatorias, debidas por ejemplo a la forma de hacer la pregunta y de contestarla.

Una solución psicométrica que integrara la noción de validez de contenido (necesaria en el contexto escolar) no nos permitiría conocer más allá de la variancia entre los individuos. Se hace necesario, por tanto, acudir a la teoría de la generalizabilidad, que permite conocer la existencia de fuentes de variancia múltiples para unificar el tratamiento de todos esos casos diferentes. En la comparación con un diseño multivariable de medidas repetidas podemos valorar el principio de simetría de las facetas (o variables) y ver la integración de la teoría de la generalizabilidad en el mismo.

Cronbach, Gleser, Nanda & Rajaratnam (1972), han desarrollado la teoría de la generalizabilidad, asumiendo que hay otras fuentes de variación además de las diferencias individuales e integrando cada una de estas fuentes de variación en una estructura global, que permite aplicaciones particulares de la teoría estadística del muestreo. La teoría de la generalizabilidad reconoce explícitamente las múltiples fuentes de error de medida (individuos, observadores, sesiones, ocasiones de medida...). Podemos estimar cada una de estas fuentes de error así como las diferentes interacciones entre ellas. El error de medida no es más que el efecto de las fluctuaciones debidas a la elección aleatoria de individuos, observadores, sesiones, ocasiones de medida..., es decir al muestreo de niveles particulares en cada una de las facetas (variables) del universo de observaciones posibles. Optimizar dicha medida es adaptar nuestro diseño para reducir al máximo la variancia del muestreo debida a estas facetas.

En la ilustración que presentamos y en otras situaciones de medida frecuentemente no se trata de diferenciar individuos, sino más bien de diferenciar ocasiones de registro, clasificación según sexo, diferentes cursos escolares, observadores, categorías, grupos de individuos, etc. En palabras de Cardinet, Tourneur & Allal (1976), esto significa asumir el principio de SIMETRÍA, es decir que sucesivos objetos de medida pueden ser evaluados dentro de un mismo diseño. Mediante dicho principio, cada faceta (variable) de un diseño puede ser seleccionada como objeto de estudio y en cada análisis de generalizabilidad de esta faceta puede ser considerada como instrumento de medida o condición de evaluación en el estudio de las otras facetas. Esta sería la **diferencia** a la solución psicométrica que tan sólo diferencia individuos.

La generalizabilidad es una teoría de los errores multifaceta de una medición conductual (Cronbach *et al*, 1972). El objetivo de la teoría es desglosar, en cualquier tipo de medición, la variabilidad real de la variabilidad del error. Para que se cumpla, la teoría necesita de los componentes del análisis de la variancia; las variaciones de las facetas (variables) tales como individuos, sesiones, ocasiones de medida, observadores, etc. El eje central de la teoría de la generalizabilidad, por tanto, se encuentra en los componentes de variancia, dado que su magnitud nos aporta información sobre las fuentes de error que están afectando una medición conductual.

La filosofía básica que subyace a la teoría de la generalizabilidad es que "un investigador se pregunta acerca de la precisión o fiabilidad de una medición dado que desea generalizar de observaciones reales a cualquier tipo de observaciones a las que éstas pertenezcan" (Cronbach, Rajaratnam & Gleser, 1983, p.144). En la investigación es el universo al que el investigador desea generalizar. El conjunto de condiciones de medida sobre las que el investigador generaliza es el universo de generalización. Este diferirá de acuerdo a los propósitos de la investigación. De acuerdo con el principio de SIMETRÍA, en cada modificación de un mismo diseño se necesita definir claramente el universo, especificando las condiciones de medida sobre las que intenta generalizar.

Los objetos de medida admisibles constituyen la población objeto de estudio y los instrumentos de medida (las condiciones de evaluación en terminología de Cronbach) constituyen el universo de generalización. Los primeros se sitúan en el aspecto de la **diferenciación**, ya que la variancia verdadera proviene de las diferencias entre objetos de estudio. Los segundos se sitúan en el aspecto de **instrumentación** (o generalización), puesto que las condiciones de medida son como los instrumentos o medios de esta medida. Es decir, estimación de la variancia verdadera debida a las diferencias entre los objetos de medida, y estimación de la variancia de error debida a la elección de los instrumentos utilizados en la medida, respectivamente.

Fundamentos teóricos de la estimación de la precisión

La obra de Cardinet & Tourneur (1985) nos abastece de procedimientos generales de cálculo, aunque otros métodos alternativos pueden encontrarse en la propia obra de Cronbach *et al.* (1972), en Brennan (1983) y en Marcoulides (1989). Basándonos en el trabajo de Cardinet, Touneur & Allal (1976) se puede en particular situar fácilmente las fuentes de error a controlar, a partir de un gráfico que represente todas las facetas del diseño.

Supongamos por ejemplo que los individuos (I), clasificados según el sexo (S), han sido evaluados en 4 cursos (C). De esta forma, se puede construir la Figura 1 (correspondiente al Plan de Medida 1a /I/S/-/C/), en la que las facetas de los "objetos de estudio" (Individuos y Sexo) están representadas mediante líneas horizontales y las facetas de los "instrumentos de medida" (Cursos) mediante líneas verticales.

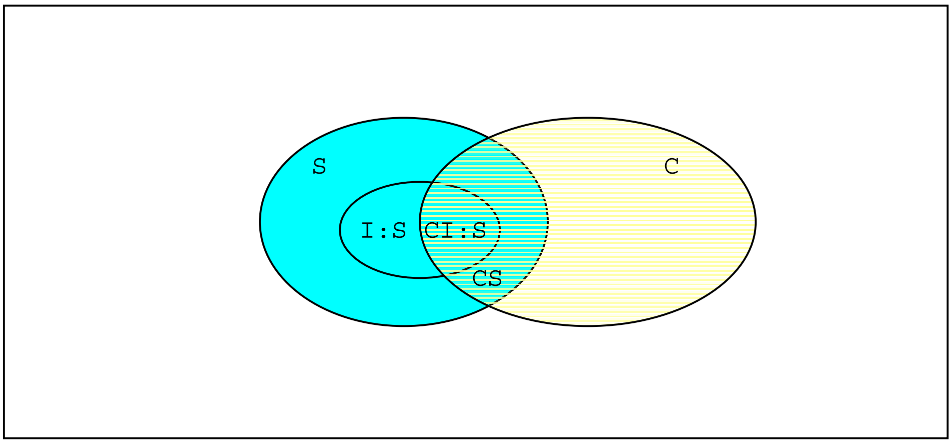

Figura 1: Representación gráfica del plan de medida 1a.

Las fuentes de error que afectan a la comparación de resultados de los individuos clasificados según el sexo se sitúan en la intersección (líneas cruzadas) de "objetos de estudio" e "instrumentos de medida". Se trata de las variancias de interacción: CS y CI:S, que el análisis de variancia permite estimar para todo el conjunto de datos correspondientes a este diseño. A partir de estas estimaciones se puede comprobar en particular si es o no más útil aumentar el número de cursos escolares.

La Figura 2 (correspondiente al Plan de Medida 5a /C/-/S/I/), en donde los "instrumentos de medida" son los individuos clasificados según el sexo, permite a partir de dichas estimaciones comprobar si es o no más útil aumentar el número de individuos.

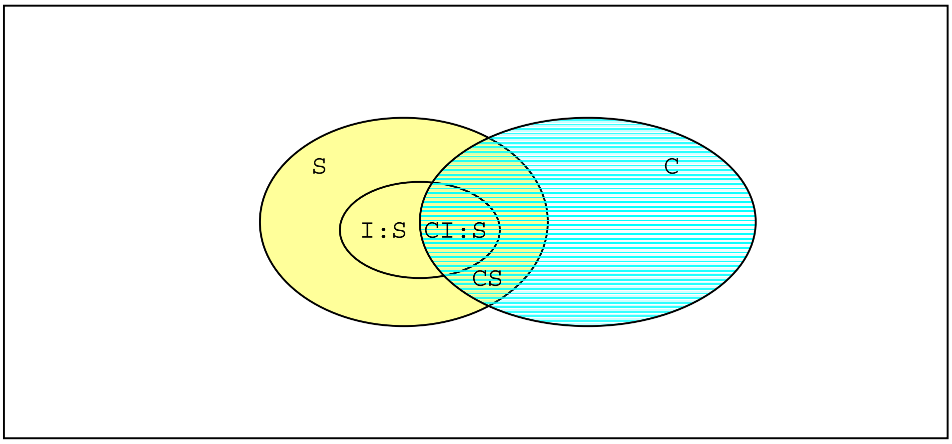

Figura 2: Representación gráfica del plan de medida 2a.

Dado el carácter abstracto, la teoría estadística puede aplicarse a cualquiera que fuese el objeto de estudio y es por ello que la teoría de la generalizabilidad constituye una teoría general de la medida (Brennan, 1983, p. 12): "el modelo de medida mejor definido actualmente existente".

La teoría de la generalizabilidad trata esencialmente la descomposición de la variancia observada en componentes de variancia y obtiene información analizando dichos componentes particularmente en lo que respecta a la contribución del error en un determinado diseño. El análisis de los componentes informa sobre qué facetas contribuyen con más error, para ser modificadas posteriormente en los sucesivos diseños. Un índice global de la variancia de puntuaciones universo relativa a la variancia del error es el coeficiente de generalizabilidad ($E\rho^2$), que se define como la proporción de variancia observada que es atribuible a la puntuación universo, es decir es la razón entre el valor esperado de la variancia de puntuaciones universo (σ^2_τ)y el valor esperado de la variancia de puntuaciones observadas ($E\sigma^2_x$)

$$E\rho^2 = \sigma^2_\tau / E\sigma^2_x \qquad (1)$$

o bien dado que $E\sigma^2_x = \sigma^2_\tau + \sigma^2_\delta$, donde σ^2_δ es la variancia de error relativo

$$E\rho^2 = \sigma^2_\tau / [\sigma^2_\tau + \sigma^2_\delta] \qquad (2)$$

o bien, dado que $E\sigma^2_x = \sigma^2_\tau + \sigma^2_\Delta$, donde σ^2_Δ es la variancia del error absoluto

$$E\rho^2 = \sigma^2_\tau / [\sigma^2_\tau + \sigma^2_\Delta] \qquad (3)$$

Podemos obtener una estimación del coeficiente de generalizabilidad (E_²) a través de estimaciones muestrales de los parámetros de la ecuación anterior

$$E\rho^2 = \sigma^2_\tau / [\sigma^2_\tau + \sigma^2_\delta] \qquad (4)$$

o bien

$$E\rho^2 = \sigma^2_\tau / [\sigma^2_\tau + \sigma^2_\Delta] \qquad (5)$$

$E\rho^2$ es sesgado, pero es un estimador consistente de $E\rho^2$ (Shavelson & Webb, 1981; Shavelson, Webb & Rowley, 1989).

La diferencia entre σ^2_δ y σ^2_Δ es que la primera no incluye las fuentes de variancia común a cada individuo (objeto de medida) mientras que la segunda sí. Por tanto, la σ_Δ^2 tendrá valores como mínimo iguales a σ^2_δ aunque casi siempre los valores serán más altos. Es interesante hacer notar que en un diseño simple, individuos por curso, $\sigma^2_i / [\sigma^2_i + \sigma^2_\delta]$ es algebraicamente (pero no conceptualmente) idéntico al α de Cronbach (Brennan, 1983).

Desarrollo del ejemplo: Planes de Medida

Para ilustrar nuestro análisis de generalizabilidad utilizaremos un diseño mixto cruzado-anidado correspondiente al plan de observación C x S (I:S). Un total de ocho individuos, clasificados según el sexo, cuatro hembras y cuatro varones, han sido evaluados en una prueba de vocabulario a lo largo de cuatro cursos escolares. Dicho plan de observación comprenderá 3 facetas, una de ellas anidada: los cursos (C, que son 4), el sexo (S, faceta necesariamente fija, igual a 2) y los individuos (I, un total de 8, anidados en el sexo, y por tanto 4 en cada sexo). Los valores figuran en una tabla de doble entrada C x S (I:S) que contiene 4 x 2 (8:2) =32 casos.

Presentamos un único diseño con los cinco planes de medida posibles, dado que al limitar la faceta sexo (S) nunca ésta podrá ser sólo faceta de generalización fija. Cada uno de ellos ha sido desdoblado en dos, para valorar el impacto que produce la transformación de la faceta, estimada aleatoriamente, cursos (C) en faceta finita N_c= 10 para los cuatro primero planes y N_c= 8 para el quinto de ellos.

En el plan 1 (ver tabla 3), el intento es el de diferenciar a los individuos anidados en la faceta sexo, ya que ellos constituirán una faceta de diferenciación. El estudio G evaluará los cuatro cursos y por tanto en diferentes ocasiones con el fin de estimar la generalizabilidad a través de los diferentes cursos. De esta forma, los cursos (C) conformarán la faceta de generalización.

En el plan 2 (ver tabla 3), se intenta diferenciar a los individuos (I), pero el investigador tiene unos objetivos diferentes dado que está interesado en la variación de los individuos independientemente de su clasificación según el sexo. Así, los individuos constituirán el objeto de estudio, mientras que el sexo y cursos serán las facetas a generalizar.

En el plan 3 (ver tabla 3), se intenta diferenciar la faceta fija sexo (S). Dado que la faceta está fijada, es evidente que el valor de la variancia de diferenciación, así como los coeficientes de generalizabilidad son nulos. Ello viene a demostrar que, en general, las facetas en clasificación y/o estratificación deberían incluirse como facetas de diferenciación con el fin de determinar si estas facetas afectan la variancia de los objetos que van a ser diferenciados (Rentz, 1987). Si los componentes de variancia de estas facetas son pequeños, deben ser eliminadas en los estudios posteriores de optimización (Cardinet, Tourneur, & Allal, 1981). Si el objeto de medida está estratificado respecto a subpoblaciones fijas, sería lógico realizar análisis por separado para cada nivel de la subpoblación además del análisis global.

En el plan 4 (ver tabla 3), los objetos de estudio serán las facetas curso (C) y sexo (S), mientras que la faceta individuos (I) constituirá la generalización. En este caso el investigador está interesado en la variación de los cursos en los dos niveles de la faceta siempre fija (S), de la que en todos los planes en que (S) ha sido diferenciación hemos presupuesto que incluye todas las posibilidades, que son dos.

En el plan 5 (ver tabla 3), se diferencian los 4 cursos (C), mientras que el objetivo del investigador es calcular la precisión de la estimación cuando quiere generalizar al conjunto aleatorio de individuos clasificados en la faceta fija sexo.

Posibles planes de estimacion y medida para ilustrar el análisis de la generalizabilidad del plan de observación C x S (I:S), N_c=4, N_s=2, $N_{I:S}$=4.

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
PLAN DE ESTIMACION	Aleatoria C,I Finita - Fija S	Aleatoria I Finita N_c=10 Fija S	Aleatoria C,I Finita - Fija S	Aleatoria I Finita N_c=10 Fija S	Aleatoria C,I Finita - Fija S	Aleatoria I Finita N_c=10 Fija S	Aleatoria C,I Finita - Fija S	Aleatoria I Finita N_c=10 Fija S	Aleatoria C,I Finita - Fija S	Aleatoria I Finita N_c=8 Fija S
PLAN DE MEDIDA	/I/S/-/C/	/I/S/-/C/	/I/-/S/C/	/I/-/S/C/	/-/S/-/C,I/	/-/S/-/C,I/	/C/S/-/I/	/C/S/-/I/	/C/-/S/I/	/C/-/S/I/
GENERALIZABILIDAD										
σ^2_τ	0.2361	0.2625	0.2361	0.2625	0.00	0.00	1.3819	1.2438	1.3264	1.1606
σ^2_Δ	0.4115	0.2469	0.3646	0.2100	0.4210	0.2028	0.1250	0.1250	0.0625	0.0625
σ^2_δ	0.0799	0.0479	0.0330	0.0198	0.0094	0.0039	0.1250	0.1250	0.3300	0.0289
COEFICIENTE DE GENERALIZABILIDAD										
RHO ABSOLUTO	$E\rho^2_{\Delta(I,S)}$ = 0.365	$E\rho^2_{\Delta(I,S)}$ =0.515	$E\rho^2_{\Delta(I)}$ =0.393	$E\rho^2_{\Delta(I)}$ =0.545	$E\rho^2_{\Delta(S)}$ =0.00	$E\rho^2_{\Delta(S)}$ =0.00	$E\rho^2_{\Delta(C,S)}$ =0.917	$E\rho^2_{\Delta(C,S)}$ =0.909	$E\rho^2_{\Delta(C)}$ =0.955	$E\rho^2_{\Delta(C)}$ =0.949
RHO RELATIVO	$E\rho^2_{\delta(I,S)}$ = 0.747	$E\rho^2_{\delta(I,S)}$ =0.846	$E\rho^2_{\delta(I)}$ =0.877	$E\rho^2_{\delta(I)}$ =0.930	$E\rho^2_{\delta(S)}$ =0.00	$E\rho^2_{\delta(S)}$ =0.00	$E\rho^2_{\delta(C,S)}$ =0.917	$E\rho^2_{\delta(C,S)}$ =0.909	$E\rho^2_{\delta(C)}$ =0.976	$E\rho^2_{\delta(C)}$ =0.976

TABLA 3

Resultados

A continuación se presentan los componentes de variancia y su contribución en porcentajes correspondientes al plan de observación original C x S (I:S), pero llevando a cabo tres posibilistas planes de estimación según nuestros objetivos (Tablas 4, 5, 6). En los tres planes de estimación mixtos no se ha variado la aleatoriedad en la que deberán ser seleccionados los individuos ni la fijación de la faceta sexo, pero si la variabilidad que podría producir la estimación aleatoria de los cursos (C) o la estimación que se produciría en el caso de poder contar con los 8 cursos escolares, dado que en nuestro ejemplo tan sólo hemos tenido en cuenta cuatro de ellos.

Tabla 4: Componentes de variancia y su contribución en porcentajes, correspondientes al plan de observación C*S(I:S), siendo N_c=$N_{i:s}$= Infinito, N_s=2

FUENTES DE VARIACION	ESTIMACION COMPONENTES DE VARIANCIA MIXTA	%
CURSO (C)	13.264	7.048
SEXO (S)	-O.0477	0
CS	556	295
I:S (INDIVIDUOS)	2.361	1.255
CI:S	2.639	1.402

Tabla 5: Componentes de variancia y su contribución en porcentajes, correspondientes al plan de observación C*S(I:S), siendo N_c=10, $N_{i:s}$= Infinito, N_s=2

FUENTES DE VARIACION	ESTIMACION COMPONENTES DE VARIANCIA MIXTA	%
CURSO (C)	11.938	6.846
SEXO (S)	-0.0422	0
CS	500	287
I:S (INDIVIDUOS)	2.625	1.505
CI:S	2.375	1.362

Tabla 6: Componentes de variancia y su contribución en porcentajes, correspondientes al plan de observación C*S(I:S), siendo N_c=8, $N_{i:s}$= Infinito, N_s=2

FUENTES DE VARIACION	ESTIMACION COMPONENTES DE VARIANCIA MIXTA	%
CURSO (C)	11.606	6.790
SEXO (S)	0.0408	0
CS	486	284
I:S (INDIVIDUOS)	2.691	1.574
CI:S	2.309	1.351

El análisis de los componentes nos aporta información sobre el diseño de medida. De hecho, el objetivo de un análisis de generalizabilidad se centra más bien en el análisis de los componentes que en los coeficientes de generalizabilidad. En el primer caso (tabla 4) N_c= infinito, se nos ofrece un componente de los individuos (I:S) relativamente pequeño, lo que no es muy deseable si los individuos constituyeran la faceta de diferenciación. Este valor aumenta a medida que disminuye la estimación del número de cursos escolares, en los otros dos planes de estimación (tablas 5 y 6), lo que afectaría a la generalización de los cursos, en el caso de que la faceta cursos (C) fuera el instrumento de medida. Es decir, sería más fácil generalizar a un número menor de cursos escolares.

En cuanto a los demás componentes de variancia, las interacciones de cursos con sexo y de cursos con individuos son también relativamente pequeñas, mientras que el componentes cursos (C) es relativamente grande, en los tres planes de estimación (tablas 4, 5, 6). Este hecho nos sugiere que, si los individuos constituyeran en el plan de medida una faceta de generalización, en estudios posteriores podríamos reducir el número de individuos, sin que por ello reduzcamos la precisión en la generalización. Por el contrario, dado que los cursos (C) contribuyen al error de una manera significativa, en el caso de que los cursos fueran la faceta de generalización en el plan de medida, necesitaríamos aumentar el número de los mismos si deseamos un nivel alto de generalizabilidad. La faceta fija sexo (S) ofrece valores nulos en los tres casos, lógico en cierta medida, ya que son muy pocos los individuos clasificados en cada nivel de esta faceta. El resultado es excelente peso no aporta mucha información a este diseño puesto que en ningún caso podrá ser considerada sola ella como faceta de generalización.

La estimación de este último componente nos ofrece un valor negativo y, aún cuando algebraicamente es imposible encontrar una suma de cuadrados negativa, es debido a las fluctuaciones muestrales en la estimación de los cuadrados medios. En estos casos, tanto Cronbach *et al.* (1972) como Cardinet *et al.* (1981), proponen reemplazar el valor negativo encontrado por un valor nulo. La modificación y optimización de este diseño, para lograr mayor precisión en la generalización de los diferentes planes de medida es lo que será considerado a continuación.

Los coeficientes absolutos [$E\rho^2_{\Delta}$] y relativos [$E\rho^2_{\delta}$] de los cinco planes de medida, desdoblados cada uno de ellos en dos, se presentan al final de la tabla 3. Estos coeficientes estiman la generalizabilidad a través de las facetas de generalización (cursos en el plan 1, sexo y cursos en el plan 2, cursos e individuos en el plan 3,...) y por ejemplo en el plan 5a, que el coeficiente de generalizabilidad tiene un valor próximo a la unidad, podría interpretarse como la correlación entre el conjunto de 100 puntuaciones y otro conjunto de otras 100 puntuaciones obtenidas en otro conjunto de 8 individuos, anidados cuatro de ellos en cada uno de los dos niveles de la faceta sexo (S). La magnitud del coeficiente (que varía de 0 a 1) se interpreta de la misma forma que los coeficientes de fiabilidad tradicionales.

Se verifica evidentemente que las medidas relativas son más generalizables que las medidas absolutas; de hecho su ambición es menos pretenciosa. En cuanto a la significación de estos resultados, podemos comprobar que las generalizabilidades no son buenas en los planes en que los individuos (I) y sexo (S) constituyen el objeto de estudio (planes 1, 2 y 3), es decir cuando el instrumento de medida que es la faceta a generalizar son los cursos. Tan solo en el plan 2b en que se lleva a cabo una estimación con un máximo de 10 cursos el coeficiente relativo tiende a acercase a la unidad. Por el contrario, si los individuos constituyen la generalización del estudio (planes 4a y 5a) obtenemos excelentes generalizabilidades en dichas situaciones de medida, ya sean absolutas ya sean relativas.

También ocurre lo mismo si transformamos la faceta de diferenciación cursos (C) de aleatoria infinita a aleatoria finita (planes 4b y 5b). Este hecho nos sugiere que los individuos, extraídos aleatoriamente de dicha población, han sido suficientes en esta disposición de medida.

A destacar, por tanto, que los coeficientes de generalizabilidad se refieren al universo de generalización. Los mayores beneficios de un análisis de generalización se derivan esencialmente de los resultados obtenidos en esta fase, que permiten llevar a cabo modificaciones en el diseño de medida y elegir un diseño optimizado (óptimo en el sentido de que se busca una máxima generalizabilidad dentro de los costos u otras restricciones prácticas o, alternativamente, que se reduzcan los costos mientras se mantenga un elevado o aceptable nivel de generalizabilidad). A continuación ilustramos algunas posibilistas modificaciones al plan de observación original.

Optimización del diseño inicial

La tabla 7 presenta el plan de medida 5a /C/-/S/I/, pero modificando sucesivamente el plan de observación original para lograr una optimización de cada una de las facetas en combinación con las otras facetas y así obtener una precisión en la generalización adecuada a este tipo de investigaciones. El hecho de limitar la población de los cursos (C), fijando la faceta cursos, no parece tener mayor incidencia en la generalizabilidad, dado que los cursos constituyen la diferenciación y no es a ellos a quien tenemos que generalizar. Conseguimos, eso sí, una mayor precisión aumentando el número de individuos de 10 a 999, en que el coeficiente de generalizabilidad valdría la unidad. Ahora bien, los costos de esta modificación serían altísimos, ya que no conseguiríamos aumentar mucho la precisión, pues en el diseño original el coeficiente relativo es 0.976. Con N_i =10 el valor sería 0.99, pero el diseño contendría 80 registros; con N_i =20 el valor sería 0.995, pero el diseño contendría 160 registros. Evidentemente la relación costo beneficio es nula, por lo que no sería aconsejable llevar a cabo modificaciones en el diseño original, ya que en consecuencia aumentarían los costos reales de la investigación.

Tabla 7: Optimización del plan 5a /C/-/S/I/.

FACETAS	NIVELES OBSERVADOS	NIVELES ESTIMADOS	OPTIMIZACION PLAN			
C	$n_c = 4$	N_c= Inf	4	4	4	4
S	$n_s = 2$	N_s= 2	2	2	2	2
I:S	$n_i = 4$	N_i= Inf	10	20	50	999
Ep²rel	976		.990	.995	.998	1.00
Ep²abs	955		.982	.991	.996	1.00

El plan de medida 5b /C/-/S/I/ (Tabla 8), exactamente igual que el anterior, pero la estimación de la faceta de diferenciación cursos (C) no ha sido infinita, sino más realista y por tanto llevando a cabo una estimación N_c = 8 que es el número de cursos escolares que constituyen nuestra educación secundaria. Los diferentes planes de optimización nos llevan a idénticos resultados. Si el valor de precisión en la generalizabilidad relativa es de 0.976 conseguiríamos aumentar a 0.99 si dispusiéramos de 20 individuos por curso en lugar de los 4 del plan original, pero en consecuencia necesitaríamos 320 observaciones; luego el precio que se paga por esta modificación es altísimo y la relación costo-beneficio nula, pues con 4 individuos por nivel casi conseguimos los mismos resultados.

Tabla 8: Optimización del plan 5b /C/-/S/I/.

FACETAS	NIVELES OBSERVADOS	NIVELES ESTIMADOS	OPTIMIZACION PLAN
C	$n_c = 4$	$N_c= 8$	8 8 8 8
S	$n_s = 2$	$N_s= 2$	2 2 2 2
I:S	$n_i = 4$	$N_i=$ Inf	20 50 75 100
E-²rel	976		.995 .998 .999 .999
E-²abs	949		.989 .996 .997 .998

En la optimización del plan 1b (tabla 9) /I/S/-/C/ obteníamos un coeficiente de generalizabilidad relativo moderadamente inferior (aunque no pequeño, 0.846), lo que indica poca precisión en la generalización de los cursos cuando el objeto de medida (puntuación universo) son los individuos anidados en la faceta sexo. Podemos apreciar que no es el número de individuos el que hace aumentar las generalizabilidades, pues aumentando de 30 a 999 el valor es idéntico. En este caso, es el número de cursos escolares (faceta de generalización) quien logra aumentar la precisión de 0.846 a 0.97 si hubiéramos dispuesto de 8 cursos escolares en lugar de los 4 del plan original. Evidentemente, las consecuencias son notables y la relación costo-beneficio bastante aceptable, dado que esta modificación, aunque con costos altísimos, permitiría una precisa diferenciación de los individuos. La prueba más evidente es la generalizabilidad absoluta del diseño original con un valor de 0.515.

Tabla 9: Optimización del plan 1b /I/S/-/C/.

FACETAS	NIVELES OBSERVADOS	NIVELES ESTIMADOS	OPTIMIZACION PLAN
C	$n_c = 4$	$N_c= 10$	8 8 8 4
S	$n_s = 2$	$N_s= 2$	2 2 2 2
I:S	$n_i = 4$	$N_i=$ Inf	30 100 999 999
E-²rel	846		.970 .970 .970 .846
E-²abs	515		.864 .864 .864 .515

Se presentan a continuación las modificaciones para optimizar al plan 4b (tabla 10) /C/S/-/I/. Podemos apreciar resultados similares a los anteriores. Aumentar el número de individuos ofrece una relación costo-beneficio bastante pequeña, pues implica tan solo aumentar el número de observaciones de 32 (en el plan original) a 2000 (en el caso $N_i = 100$), y con unos costos altísimos. Obtenemos excelentes valores de generalizabilidad considerando 10 cursos escolares y 10 individuos por cada nivel de la faceta sexo (0.961).

Tabla 10: Optimización del plan 4b /C/S/-/I/.

FACETAS	NIVELES OBSERVADOS	NIVELES ESTIMADOS	OPTIMIZACION PLAN
C	$n_c = 4$	$N_c = 10$	10 10 10 10
S	$n_s = 2$	$N_s = 2$	2 2 2 2
I:S	$n_i = 4$	$N_i = Inf$	10 50 75 100
E-²rel	909		.961 .992 .995 .996
E-²abs	909		.961 .992 .995 .996

Discusión de resultados

Concluimos, por tanto, que no serían suficientes las modificaciones al plan de estimación (ya pusimos de manifiesto el elevado componente de variancia de cursos y su alta contribución al error, 1.32 y 70.48, respectivamente en el plan de estimación aleatoriamente infinito) y se necesitaría una revisión de la faceta cursos para plantear con éxito investigaciones futuras. En cuanto a las anidaciones que se presuponen en esta ilustración, que reducen los costos en el número de niveles de las facetas, determinan una buena precisión y no hacen aumentar el número de niveles que nosotros habíamos restringido con la anidación. El análisis de generalizabilidad ha servido de estudio piloto para plantear de nuevo una recogida de datos más coherente al diseño de medida.

En consecuencia, el diseño inicial multivariable de medidas repetidas no nos asegura precisión en los resultados cuando han de generalizarse a un número mayor de muestras y, por tanto, o se eligen nuevos niveles para cada una de las facetas o se aumenta el número de los mismos o, finalmente, rediseñamos nuestro estudio en función de la información suministrada por la teoría de la generalizabilidad.

La ventaja de este tipo de análisis es la capacidad para diseñar estudios D más eficientemente sobre la base de la información aportada por los planes de estimación y medida. Si tenemos en cuenta la relación nivel de precisión-costos, es obvio que podremos diseñar un estudio D óptimo (Johnson & Bell, 1985).

Como conclusión general a lo que aquí se ha mostrado, la disimilitud básica entre el Método Multivariable y la Teoría de la precisión en la estimación (Generalizabilidad), es que el método multivariable es aplicado como diseño, por lo tanto se busca el efecto de las variables o tratamientos. Mientras que la Teoría de la Generalizabilidad estudia el alcance de un resultado, ya que permite adaptar el diseño a los intereses que marca la teoría en la que se encuadra el estudio.

Como semejanza básica encontramos al modelo de la regresión como estructura de análisis. La generalizabilidad es un estudio apriorístico que nos permite determinar el número de niveles de las variables que posteriormente permitirán configurar el diseño multivariable. Esto nos simplificará el desarrollo de este diseño, obteniendo los mismos o mejores resultados que si utilizamos el diseño clásico cuyo criterio es elegir arbitrariamente las variables.

Optimización de resultados y maximización del coeficiente G

Aunque es importante obtener buena precisión a través de los diferentes coeficientes de generalizabilidad, no siempre basta con la optimización de los mismos si se imponen ciertas condiciones restrictivas al diseño. Estas restricciones nos permitirán evaluar si la precisión total del diseño se ajusta al presupuesto disponible para llevar a cabo la investigación. Se nos plantea ahora el problema de maximización del coeficiente de generalizabilidad en función de un conjunto pre-especificado de limitaciones presupuestarias. Por ejemplo, en un diseño cruzado de dos facetas individuos x cursos (i x c), es bastante sencillo cumplir con las condiciones restrictivas a medida que se maximiza el coeficiente de generalizabilidad. Para ello, bastaría con elegir el mayor número de cursos que maximize el coeficiente de generalizabilidad, sin violar el presupuesto. No es así cuando se añaden otras facetas al diseño, obteniendo una solución que es algo más complicada.

De acuerdo a esta idea, Woodward & Joe (1973) investigaron el efecto de la precisión cuando se alteraba el número de observaciones de las facetas en diferentes diseños multi-facetas. Básicamente, estos autores nos ofrecen la forma de determinar el número óptimo de observaciones a realizar en orden a maximizar el coeficiente de generalizabilidad; ahora bien, siempre que sean cruzadas y aleatorias todas ellas. En el caso de dos facetas, siendo n_i y n_j el número de observaciones de las facetas i y j , presentan una fórmula cerrada para determinar n_i y n_j. Sin embargo, en el caso de tres facetas, siendo n_i, n_j y n_k el número de observaciones, presentan una solución que requiere resolver las raíces de un conjunto polinómico de orden seis. Además, aunque no consideran otros diseños multi-faceta, los diseños de más de 3 facetas requieren la resolución de las raíces de los polinomios, que aumentan el orden del polinomio en proporción al número de facetas de que consta el diseño. Por tanto, no parece una tarea asequible, para muchos investigadores que no son metodólogos y que han de tomar decisiones, la determinación del número óptimo de observaciones a realizar en un diseño en que el número de facetas es mayor que 2.

Recientemente, Marcoulides & Goldstein (1990) nos ofrecen una solución alternativa al trabajo de Woodward & Joe (1973). El procedimiento determina, también para facetas cruzadas y aleatorias, el número óptimo de condiciones cuando se imponen limitaciones presupuestarias a un diseño. De esta forma, se maximiza el coeficiente en función de los costes reales de la investigación. La elección de facetas cruzadas y aleatorias para dicha maximización nos presentaría los casos más desfavorables del diseño, ya que es donde los componentes de variancia adquieren el máximo de variabilidad. Aún así, hemos comprobado que las diferencias suelen ser mínimas y no afecta al número óptimo de condiciones que maximiza el coeficiente.

En nuestro caso, disponemos de un diseño de 3 facetas (cursos, individuos, sexo). Es evidente que para encontrar ese coeficiente de generalizabilidad máximo relativo, sin violar limitaciones presupuestarias, necesitamos seleccionar un cierto número de observaciones para cada faceta. Este problema de elección es similar al que nos presentan Woodward & Joe (1973). Ahora bien, consideramos que la optimización realizada previamente nos ofrece ya ese número máximo de observaciones para cada faceta; no conocemos sin embargo si ese número es el óptimo en función de los costes reales de la investigación. Y esa es la cuestión que vamos a tratar de determinar.

Imaginemos uno de nuestros posibles planes de medida, en el que la faceta cursos constituye la diferenciación y las facetas sexo e individuos la instrumentación. El problema es una optimización no-lineal, en el que la decisión estará referida al número de individuos (n_i) en función de la clasificación por el sexo (n_s). El objetivo de optimización es maximizar el coeficiente de generalizabilidad ($E\rho^2_\delta$), que se define como

$$E\rho^2_\delta = \frac{\sigma^2_c}{\sigma^2_c + \sigma^2_\delta} \tag{6}$$

donde

$$\sigma^2_\delta = \frac{\sigma^2_{ci}}{n_i} + \frac{\sigma^2_{cs}}{n_s} + \frac{\sigma^2_{cis}}{n_i n_s} \tag{7}$$

con la consiguiente limitación presupuestaria

$$p n_i n_s \leq \overline{p}$$

donde

p = presupuesto para la interacción individuo-sexo, y
$\overline{p}$ = presupuesto total disponible en la investigación.

Se trata de encontrar un conjunto de valores para n_i y n_s que maximice $E\rho^2$. Dado que en la optimización σ^2_c es una constante, el objetivo es minimizar σ^2_δ bajo las restricciones presupuestarias en las que $p\ n_i\ n_s \leq \overline{p}$. Para ello utilizaremos la función de Lagrange L (n_i, n_s, λ) = σ^2_δ - $\lambda(p\ n_i\ n_s - \overline{p})$ diferenciando con respecto *a* n_i, n_s y λ e igualando a cero la expresión resultante. La minimización quedaría como sigue:

$$\min F(n_i, n_0, \lambda) = \frac{\sigma^2_{ci}}{n_i} + \frac{\sigma^2_{cs}}{n_s} + \frac{\sigma^2_{cis}}{n_i n_s} - \lambda(p n_i n_s - \overline{p}) \tag{8}$$

Diferenciando con respecto a n_i, n_o , λ e igualando las derivadas a cero obtenemos:

$$n_i = \sqrt{\frac{\sigma^2_{ci}}{\sigma^2_{cs}}\left(\frac{\overline{p}}{p}\right)} \tag{9}$$

$$n_s = \sqrt{\frac{\sigma^2_{cs}}{\sigma^2_{ci}}\left(\frac{\overline{p}}{p}\right)} \tag{10}$$

Teniendo en cuenta los componentes de variancia aleatorios obtenidos en el caso de que nuestro diseño fuera totalmente cruzado (Tabla 11) e imponiendo limitaciones presupuestarias al diseño, ya que el costo total de investigación es de 192.000 ptas. y el costo por individuo clasificado según el sexo es de 4.700 ptas en cada interacción del diseño, maximizaríamos el coeficiente de generalizabilidad contando con 11 individuos en lugar de los 8 con los que se llevó a cabo la investigación. Ello supondría la mejor relación costo-beneficio del diseño sin necesidad de aumentar en gran cuantía el número de observaciones. Similares resultados se obtienen en el caso de considerar a las facetas cursos e individuos en la instrumentación: aumentando a 10 individuos y manteniendo el mismo número de cursos conseguiríamos la maximización del coeficiente, tanto en el caso de considerar a la faceta sexo fija como aleatoria. También en este caso se han limitado los presupuestos (p= 4.700 ptas y $\overline{p}$ = 192.000 ptas).

Tabla 11: Componentes de variancia aleatorios para el plan C x I x S (C=I=S= Infinito).

FUENTES DE VARIACION	VARIANCIA	%
C	11.458	6.310
S	-972	0
CS	1.181	650
I	-833	0
CI	417	229
SI	2.847	1.568
CSI	2.257	1.243

En definitiva, las conclusiones parecen ser contrarias a las obtenidas previamente en la optimización de los diferentes planes de medida. ¿Qué ha cambiado?.

Básicamente, no hemos aumentado en precisión, pero hemos establecido el número óptimo de observaciones a realizar disponiendo de 192.000 ptas para llevar a cabo la investigación y de 4.700 ptas para cada interacción.

A tenor de los resultados obtenidos en el proceso de optimización para llevar a cabo una *diferenciación* con precisión necesitaríamos de 8 cursos escolares en lugar de los 4 fijados inicialmente en el diseño, pero en contrapartida serían 4 años más de investigación y un presupuesto total multiplicando por 4 (exceptuando gastos de inflación y de crédito). Los costos serían altísimos y quizás no se llevaría a cabo la investigación. La maximización ajusta el presupuesto a la relación costo-beneficio, aunque ello suponga no aumentar en precisión. En todo caso, el dilema quedaría en manos de la toma de decisión del investigador. Es decir, conseguir una buena precisión en la generalización de sus resultados aumentado el número de cursos escolares y manteniendo casi el mismo número de individuos (optimización) o estimar con menor precisión manteniendo el mismo número de cursos escolares y aumentando tan solo a 10 el número total de individuos en la investigación pero contando con un presupuesto real de 192.000 ptas (maximización).

Por supuesto, se pueden añadir otras limitaciones adicionales a la maximización que hemos llevado a cabo; por ejemplo, podemos imponer un número máximo de cursos y un número mínimo de individuos antes de optimizar la solución. Tales restricciones adicionales pueden ser importantes cuando, a pesar de restricciones presupuestarias, el investigador ha de tomar en consideración esas otras restricciones del diseño.

Referencias

Arnau Gras, J. (1981). *Diseños experimentales en psicología y educación*. Vol. 1. México: Trillas.
Arnau Gras, J. (1990). *Diseños experimentales multivariables*. Madrid: Alianza Editorial.
Arnau Gras, J., Blanco Villaseñor, A., y Losada López, J.L. (1990, Noviembre). *Semejanzas y disimilitudes entre el método multivariable (medidas repetidas) y la teoría de estimación de la precisión (generalizabilidad).* Comunicación presentada en el VIII Congreso Nacional de Psicología (Mesa Redonda 'Métodos y Técnicas'). Barcelona.
Bartlett, M.S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, 9*, 176-197.
Blanco Villaseñor, A. (1989). Fiabilidad y generalización de la observación conductual. *Anuario de Psicología*, *43* (4), 5-32.
Bock, R.D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
Brennan, R.L. (1983). *Elements of generalizability theory*.Iowa City, Ia.: The American College Testing Program.
Cardinet, J., Tourneur, Y., & Allal, L. (1976). The simmetry of generalizability theory: Applications to educational measurement. *Journal of Educational Measurement, 13* (2), 119-135.
Cardinet, J., Tourneur, Y., & Allal, L. (1981). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement,18* (4), 183-204.
Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure*. Berne: Peter Lang.
Cardinet, J. (1987). *La construction de tests d'apprentissage selon la théorie de la généralisabilité* (Recherches; 87.107). Neuchàtel: Institut romand de recherches et de documentation pédagogiques.
Cochran, W.G. & Cox, G.M. (1957). *Experimental designs*. New York: John Wiley (edición castellana, 1974. México: Trillas).
Cronbach, L.J., Rajaratnam, N., & Gleser, G.C. (1963). Theory of generalizability: a liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology, 16*, l37-163.
Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: Wiley.
Dixon, W.J., Brown, M.B., Engelman, L., & Jennrich, R.I. (1990). *BMDP Statistical Software Manual.* Berkeley, Ca.: University of California Press.
Dayton, C.M. (1970). *The design of educational experiments*. New York: McGraw-Hill.
Fisher, R.A. (1925). *Statistical methods for research workers* (1ª ed.). Londres: Oliver and Boyd.
Federer, W.T. (1955). *Experimental design, theory and application*. New York: Macmillan.
Fleiss, J.L. (1986). *The design and analysis of clinical experiments.* New York: John Wiley.
Marcoulides, G.A. (1989). The application of generalizability analysis to observational studies. *Quality & Quantity: The International Journal of Methodology, 23*(2), 115-127.
Marcoulides, G.A., & Goldstein, Z. (1990). The optimization of generalizability studies with resource constraints. *Educational and Psychological Measurement, 50*, 761-768.
Rentz, J.O. (1987). Generalizability Theory: A Comprehensive method for assessing and improving the dependability of marketing measures. *Journal of Marketing Research, 24*, 19-28.
Shavelson, R.J., & Webb, N.M. (1981). Generalizability theory: 1973-1980. *British Journal of Mathematical and Statistical Psychology, 34*, 133-166.
Shavelson, R.J., Webb, N.M., & Rowley, G.L. (1989). Generalizability Theory. *American Psychologist, 44* (6), 922-932.
Tukey, J.W. (1949). Dyadic ANOVA, an analysis of variance for vectors. *Human Biology, 21*, 65-110.
Wilks, S.S. (1932). Certain generalization in the analysis of variance. *Biometrika, 24*, 471-494.
Woodward, J.A., & Joe, G.W. (1973). Maximizing the coefficient of generalizability in multi-facet decision studies. *Psychometrika, 38*(2), 173-181.

Original recibido: 29-4-91
Aceptado: 7-6-91