

ALGUNAS CONSIDERACIONES RELATIVAS AL ANÁLISIS DE DISEÑOS EXPERIMENTALES MULTIVARIABLES

G. VALLEJO SECO; F. J. HERRERO DÍEZ

Universidad de Oviedo

Resumen

En la investigación psicológica actual un fenómeno cada día más constatado es la creciente utilización de los diseños experimentales con múltiples variables dependientes. Y si bien es cierto que durante décadas este hecho fue poco común dentro de la estrategia de investigación de carácter más experimental, no lo es menos que hoy en día, por causa de su capacidad para manejar aspectos teóricos, metodológicos e inclusive estadísticos, goza de mayor notoriedad. Con todo, el registrar diversas medidas desde una misma unidad de análisis no está exento de problemas; por este motivo, en el presente trabajo hemos centrado nuestra investigación en tres aspectos que, a nuestro modo de ver, son fundamentales, a saber: a) ¿por qué manejar dos o más variables dependientes simultáneamente?, b) ¿cuántas variables dependientes emplear?, y c) ¿cuál de los varios procedimientos analíticos disponibles utilizar?

Introducción

Durante los últimos tiempos, un fenómeno cada vez más constatado es la creciente utilización, por parte de los experimentadores, de diseños de investigación con múltiples variables dependientes registradas desde una misma unidad de análisis. Así, por ejemplo, un investigador clínico puede estar interesado en evaluar el efecto de la implantación de un programa basado en la técnica de la economía de fichas sobre la reducción de la depresión neurótica, consumo de drogas y disputas familiares de un grupo de pacientes; un investigador en el área de la psicología escolar puede desear conocer los efectos de un programa educativo de televisión sobre la conducta cooperativa, motivacional y actitudinal de un grupo de niños de corta edad, y un investigador en el campo de la psicología animal puede buscar profundizar en la relación existente entre los efectos conjuntos de la demora de la recompensa, la canti-

Abstract

Nowdays it is becoming very frequent the use of Multiple Experimental Designs which include several dependent variables measured on the same subject. It is true that this fact is not a new one in Psychology, in fact, during decades it has been used almost exclusively in the framework of investigations based in selective manipulation; but, currently, it must be noticed that this type of analysis plays a relevant role in those investigations based on any kind of manipulative techniques. In this paper we have focused our research in three main aspects: a) Why to use more than one dependent variable simultaneously; b) How many dependent variables must be used, and c) Which one of the possible analytic procedures is more suitable to be employed.

Finally, the implications of these three points for the future psychological research are discussed.

dad de reforzamiento y el esfuerzo requerido en la ejecución de las respuestas, por un lado, y la velocidad de recorrido, resistencia a la extinción y la recuperación espontánea, por el otro. En todos los casos el investigador registra múltiples variables dependientes.

Y, si bien es cierto que este hecho no constituye por sí mismo ninguna novedad dentro de la psicología, ya que durante décadas ha sido utilizado, casi de un modo exclusivo, dentro del enfoque de investigación basado en la manipulación selectiva (perspectiva psicométrica-diferencial), no por ello debemos dejar pasar por alto que hoy en día ocupa un lugar destacado en la estrategia de investigación basada en la manipulación experimental de cualquier signo. Para ilustrar lo dicho, baste apuntar que Zinkgraf (1983) refiere tres estudios de carácter estadístico-bibliométrico en las áreas de clínica, experimental y evolutiva, donde se pone de relieve que actualmente, dentro de estas materias, se utilizan

diseños con más de una variable de medida en más del 50 por 100 de las veces.

Ante lo dicho, se nos ocurren tres preguntas de inmediato: ¿por qué los investigadores están cada día más interesados en evaluar si existen diferencias entre las medias de diferentes grupos en dos o más variables dependientes comparadas simultáneamente, en vez de analizar una cada vez?, ¿cuántas variables dependientes debemos utilizar? y ¿cuál de las diferentes estrategias disponibles para analizar medidas múltiples hemos de utilizar?

¿Por qué utilizar múltiples variables dependientes?

Por lo que respecta a la cuestión inicial, existen al menos tres razones importantes por las que un investigador podría estar interesado en registrar más de una variable de medida. En primer lugar, consideraciones de tipo teórico. Evidentemente, dado que el sistema psicológico constituye un todo integrado, es de esperar que cualquier tratamiento afecte al sujeto de forma variada, no sólo en un sentido; por tanto, considerar más de un criterio de medida nos permite una descripción más detallada del fenómeno que tenemos bajo investigación. Por ejemplo, si dos o más tratamientos para reducir la ansiedad fueran sometidos a prueba resultaría mucho más informativo registrar la ansiedad por procedimientos diversos (Escalas de ansiedad de Taylor, procedimientos psicofisiológicos, etc.) que mediante un solo registro.

En segundo lugar, apreciaciones de tipo metodológico. La utilización de varios criterios de medida nos permite generar información más «válida», tomando esta palabra en el sentido de Cook y Campbell (1979). Así pues, multioperacionalizar el constructo o registrar varias conductas nos va a permitir incrementar la validez de nuestras investigaciones. En efecto, en las ciencias sociales y comportamentales donde las medidas son, en general, menos perfectas (lo cual atenta directamente contra la validez de conclusión estadística, al incrementarse la varianza de error), tener varios criterios y utilizar métodos multivariados para evaluar el efecto de los tratamientos puede proporcionarnos una información más precisa de la relación causal existente entre los antecedentes y los consecuentes (Cole, Howard y Maxwell, 1981). Además, el análisis simultáneo del efecto de los tratamientos sobre múltiples variables dependientes se convierte en un procedimiento de primera magnitud al permitirnos indagar acerca de la validez de constructo (Judd y Kenny, 1981), tanto en su vertiente convergente (grado mediante el cual las variaciones en las medidas son el resultado de variaciones en el constructo teórico de interés), como en su vertiente discriminante (grado por el cual las variaciones en las medidas no lo hacen como resultado de las variaciones de otros constructos que carecen de interés). Este hecho es consecuencia directa de la posibilidad que tenemos de

estimar las correlaciones entre los indicadores y el constructo desde las correlaciones observadas entre los indicadores (Kenny, 1979).

Y, por último, puntualizaciones de tipo estadístico. En efecto, si, como ocurre frecuentemente, un investigador está interesado en indagar acerca de la relación causal entre dos grupos de variables, la estrategia univariada, de efectuar *análisis univariados de la varianza* (AVAR), o de la regresión múltiple separadamente con cada una de las variables dependientes, acarrea una serie de problemas en relación con la fijación exacta del nivel de significación (α) a la hora de efectuar decisiones estadísticas (Miller, 1981), con la interpretación de los resultados y con la potencia de la prueba. A continuación comentaremos brevemente cada uno de estos inconvenientes. La utilización de pruebas de significación múltiples incrementa la probabilidad de extraer una o más conclusiones erróneas como una función del número de pruebas ejecutadas, a menos que α haya sido adecuadamente ajustada, o se prueben las hipótesis acerca de todas las variables de medida simultáneamente. En efecto, uno de los inconvenientes que conlleva el efectuar comparaciones múltiples de forma independiente cuando en realidad no lo son, sobre todo cuando se realiza una cantidad de comparaciones elevadas a un mismo nivel de α , es la probabilidad cada vez más alta de cometer errores de Tipo I (*the experimentwise error rate*). La probabilidad de cometer un error de Tipo I en una sola comparación nos viene dada por el α establecido. De este modo, si un investigador lleva a cabo C pruebas separadamente, cada una con un nivel α la probabilidad de cometer un error de Tipo I es $1 - (1 - \alpha)^C$, lo cual es aproximadamente igual a $C\alpha$ para pequeños valores de α . Por ende, $(1 - \alpha)^C$ constituye la probabilidad de quedarse con las hipótesis de nulidad resultantes de las comparaciones cuando son verdaderas. Por ejemplo, si efectuamos 8 pruebas F, cada una a un $\alpha = 0,05$, la probabilidad de rechazar al menos una de ellas falsamente es aproximadamente 0,34; lo que resulta del todo inaceptable.

Igualmente, llevar a cabo pruebas univariadas separadas puede causar problemas interpretativos en el caso de que las variables dependientes estén intercorrelacionadas. Por ejemplo, imaginémosnos que un investigador encuentra un efecto positivo de su tratamiento en la reducción de conductas no deseadas a las cuales subyace el modelo causal esquematizado en la figura 1.

¿Cuáles serían en este caso las conclusiones del investigador?, ¿hubiesen sido las mismas de no haberse mantenido ningún tipo de correlación entre las variables, como aparece en la figura 2?

Obviamente, las dos situaciones son distintas, los análisis requeridos también son distintos (o al menos deberían serlo) y las conclusiones a las que se llegaría acerca de la implementación del tratamiento también lo serían.

Otro problema que se nos puede presentar al efectuar pruebas univariadas una cada vez tiene que ver con el hecho de no encontrar diferencias

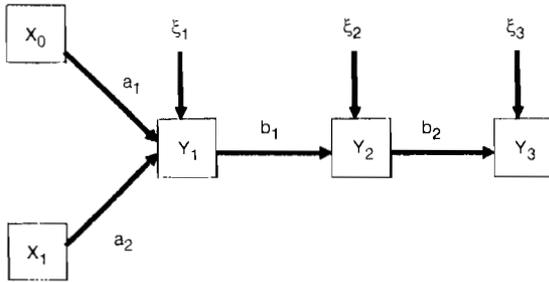


Figura 1. Diagrama causal de efectos indirectos.

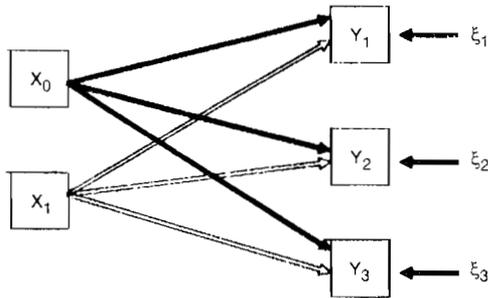


Figura 2. Diagrama de causas independientes y efectos directos.

entre los tratamientos al considerar cada variable dependiente individualmente, mientras que si consideramos todas las variables simultáneamente puede ocurrir que fuéramos capaces de obtener diferencias significativas. De ahí que hayamos manifestado más arriba que las pruebas multivariadas son, generalmente, más poderosas que las univariadas. Aunque la potencia de las pruebas multivariadas es un tema sometido a constante debate (Bray y Maxwell, 1985), sin embargo, lo común es que dependa de la adecuada elección del tamaño de muestra, del número de variables dependientes a incluir en el análisis, del tamaño del efecto que se desea conseguir y de la dirección y magnitud de las intercorrelaciones dentro de los grupos (Muller y Peterson, 1984). En consecuencia, lo dicho en este punto, así como la consideración efectuada más arriba, debe hacernos reflexionar acerca de la sugerencia harto frecuente de efectuar análisis univariados independientes si la hipótesis nula multivariada ha sido rechazada. Pues, como acertadamente señala Pedhazur (1982), esta sugerencia es una pésima recomendación, ya que, al ignorar las intercorrelaciones entre las variables, desechamos la información que las pruebas multivariadas recogen a través de la matriz de covarianza, apartándonos de este modo del verdadero propósito del análisis multivariado. Discusiones teóricas e ilustraciones gráficas sobre esta cuestión las podemos encontrar en Li (1977: 443-444), Tatsuoka (1971: 22-24),

Timm (1975: 166) y Bray y Maxwell (1985: 11-12 y 30-31).

Finalmente, no deseamos terminar este apartado pasando por alto las posibles desventajas a las que conduce la utilización del análisis multivariado, dentro del contexto en el que nos movemos; bien sea por razones de potencia, bien sea por razones de complejidad, o por ambas a la vez. Por ejemplo, en el supuesto caso de disponer de múltiples variables dependientes, inclusive con una alta correlación intraclassa (0,60-0,80), no es inverosímil que ocurra que la hipótesis nula multivariada probada mediante un análisis multivariado de la varianza (AMVAR) sea aceptada con un nivel de significación del 10 por 100; mientras que si llevásemos a cabo análisis de la varianza separados con cada una de las medidas nos podríamos encontrar con la enorme sorpresa de que todas fueran significativas al 5 por 100. Situaciones como esta, en que la potencia de prueba está muy cuestionada, pueden ocurrir con más frecuencia de lo deseado, al aplicar pruebas multivariadas con un escaso número de tratamientos (Harris, 1985) y/o en el caso de disponer de pequeños o moderados tamaños de muestra (Bray y Maxwell, 1985). Sin olvidar que este hecho se acentúa bastante más en las ocasiones en que las variables sean estadísticamente independientes y que se efectúe un AMVAR.

Por lo demás, tampoco quisiéramos dejar sin un breve comentario el siguiente aspecto: con relativa frecuencia, un investigador antes de registrar sus datos planifica ejecutar una prueba estadística multivariada, comúnmente, un AMVAR, con el único fin de controlar la tasa de error por experimento. En el caso de cumplirse los supuestos que requieren aplicar este tipo de pruebas, la situación aparentemente es correcta; de manera que cuando éste es el caso, la combinación de una prueba multivariada significativa con pruebas F univariadas ha sido frecuentemente recomendada por Finn (1974), Bock (1975), Wilkinson (1975), Hummel y Sligo (1971) para interpretar las diferencias entre los grupos. Sin embargo, como ha sido advertido por varios autores (Timm, 1975; Strahan, 1982; Bray y Maxwell, 1982), esta forma de analizar las variables dependientes adolece de algunas desventajas que nos deben hacer recapacitar sobre su complementación. En concreto, si bien la primera parte de la prueba nos protege de la posibilidad de cometer error de Tipo I, no ocurre lo mismo en el caso de que la hipótesis nula global sea rechazada por ser sólo parcialmente verdadera. Es decir, que en el caso de disponer de dos o más variables dependientes, si rechazamos la hipótesis nula multivariada al 5 por 100, porque al menos una de las variables produce diferencias entre los tratamientos de la población, no tenemos garantías suficientes de que el nivel de significación se mantenga protegido a la hora de probar la parte de la hipótesis nula que aún permanece como verdadera.

Así pues, en el caso de que el propósito de nuestro análisis lo constituya el controlar la tasa de error ante la ejecución de pruebas de significación múltiples podemos optar por efectuar análisis de la va-

rianza o regresión con el nivel α ajustado mediante la desigualdad de Bonferroni (α /número de variables dependientes) y comparar cada razón F teórica con el nivel α corregido, o bien, como señalan Huberty y Morris (1989), optar por el procedimiento de prueba simultáneo inventado por Gabriel (1969), propugnado por Bird y Hadzi-Pavlovic (1983), y programado por O'Grady (1986). Conviene recordar que estos procedimientos tampoco están libres de problemas, pues mientras el primero de los dos procedimientos atenúa la potencia de prueba a medida que el número de variables se incrementa, el segundo es en exceso conservador por su capacidad para protegernos a un mismo nivel α para todas las posibles comparaciones que realicemos. Ahora bien, en el contexto en el que nos movemos, si bien no nos sería difícil disponer de un gran número de variables, sí nos sería difícil lograr que las variables estuviesen libres de correlación, que ha sido el aspecto en torno al que hemos pretendido situar esta solución. Es decir, que esta salida sería recomendable, sobremedida, cuando deseamos averiguar la contribución de los diferentes tratamientos en cada una de las variables y éstas sean conceptualmente independientes, o, al margen de la posible interrelación, al investigador le interesa conducirse así por alguna razón especial, como puede ser la naturaleza exploratoria de la investigación, o bien la necesidad de explicar algún resultado previamente obtenido en aras a su posible generalización.

¿Cuántas variables dependientes utilizar?

Básicamente, hasta ahora, nos hemos estado refiriendo a las excelencias de evaluar si existen diferencias entre las medias de dos o más tratamientos en dos o más variables dependientes comparadas simultáneamente. En concreto, hemos dicho que el incluir múltiples variables dependientes en un diseño nos capacita para descubrir relaciones complejas entre las variables, para comprender mejor el fenómeno bajo investigación, dado que la ejecución de muchas tareas no es fácilmente representada por una sola variable, para protegernos mejor frente a los errores de Tipo I, e inclusive de Tipo II y, en definitiva, para incrementar la validez de nuestras investigaciones. Con todo, aunque el enfoque multivariado nos permite manejar un gran número de variables de medida, éstas deben ser cuidadosamente seleccionadas, si queremos preservar las ventajas mencionadas. Pues, si no, puede ocurrir que, al conducir una investigación donde las variables fueran registradas sin seguir ningún criterio teórico y/o empírico, no se manifiestan como tales tratamientos significativos. De ahí que sea determinante elegir adecuadamente las variables. Seguidamente, basándonos fundamentalmente en los trabajos de Stevens (1986) y de Bray y Maxwell (1985), vamos a especificar una serie de considera-

ciones cuya observación nos proporciona una estimable ayuda frente al uso indiscriminado de múltiples variables dependientes:

1. Las variables no deberían ser incluidas en el análisis solamente por estar disponibles, pues un reducido número de ellas, además de añadir robustez, facilita la interpretación de los resultados (Olson, 1974).
2. Al incluir variables de medida sin una base racional puede ocurrir que pequeñas diferencias en algunas de ellas oscurezcan las diferencias reales existentes en otras. Las pruebas multivariadas detectan principalmente errores en el sistema (p. ej., en las variables), y por esta razón podemos aceptar erróneamente la hipótesis de nulidad multivariada.
3. La potencia de las pruebas multivariadas disminuye a medida que el número de variables dependientes aumenta (Das Gupta y Perlman, 1973). De aquí que sea sensato limitar el número de variables, a no ser que incrementemos el tamaño de muestra, disminuyamos el error de medida y las variables estén elegidas en base a criterios teóricos y/o empíricos.
4. En áreas donde la medida no está lo suficientemente desarrollada, como ocurre frecuentemente en las ciencias sociales, comportamentales y de la salud, el uso de criterios múltiples puede mostrarse problemático. Por tanto, en estos casos, es conveniente tener presente la recomendación hecha por Pruzek (1974), en el sentido de combinar linealmente las variables de medida altamente similares, sobre todo, cuando la fiabilidad e intercorrelaciones de estos criterios difieren sustancialmente. En concreto, técnicas de reducción de datos, tipo análisis de componentes principales, podrían ser empleadas como un paso preliminar, ya que siempre se debería considerar la posibilidad de que un reducido número de factores dé cuenta de la mayor parte de la varianza contenida en las variables originales. Con ello se intenta que el fenómeno bajo estudio sea presentado de una forma tan simple como sea posible, pero sin perder información valiosa.

¿Cuál de los diferentes procedimientos analíticos utilizar?

Dar una respuesta concreta acerca de cuál de las diferentes estrategias analíticas utilizar representa, sin duda, una tarea de gran dificultad. Pues, si bien es cierto que cada día son más frecuentes las publicaciones en las que los experimentadores están interesados en evaluar los efectos de los tratamientos en múltiples variables dependientes, como se des-

prende de los actuales estudios dentro de las áreas clínica, experimental y evolutiva, sin embargo, no lo es menos que la forma de hacerlo difiere considerablemente.

Dentro de este amplio abanico de publicaciones hemos de destacar sobre todo las que utilizan los diseños de análisis de la varianza con múltiples variables de medida, complementados, frecuentemente, con otros enfoques, en el supuesto de que la hipótesis nula multivariada sea rechazada. Enfoques que, a nuestro modo de ver, pueden ser caracterizados de forma diferente en función de que el investigador desee analizar preferentemente cuestiones relacionadas con la importancia de los grupos implicados en el análisis (contrastes univariados o multivariados), cuestiones referidas a la trascendencia de las diferentes variables dependientes (análisis step-down, pruebas F separadas, análisis discriminante, etc.), o bien cuestiones concernientes a la relevancia de las variables, tanto independientes como dependientes de modo simultáneo. Con todo, tampoco es infrecuente encontrarse con trabajos en los que los autores hacen uso de las técnicas tradicionalmente ligadas a la práctica correlacional, dentro de un contexto, básicamente, experimental. Técnicas tales como el análisis factorial confirmatorio, el análisis factorial canónico, o los modelos de estructura de covarianza en los que se asume que las variables independientes afectan a las variables latentes también han sido ocasionalmente utilizados en investigaciones de corte experimental, como ponen de relieve los trabajos de Alvin y Tessler (1974), Bagozzi, Fornell y Larcker (1981) o Palmer, Drummond, Tollison y Zinkgraf (1984). Aspecto éste que no nos debe sorprender, pues es un hecho bien conocido que lo que hace que una investigación sea o no experimental no depende de las técnicas analíticas utilizadas, sino de los procedimientos de manipulación y de control empleados.

A la vista de estas dos formas de proceder, y aun a riesgo de ser en exceso sintéticos, podemos decir que, en el primer caso, los análisis estadísticos estarían primordialmente determinados por el diseño experimental que se utiliza, mientras que, en el segundo, la estrategia analítica a seguir estaría condicionada tanto por el diseño como por el modelo causal que subyace a las variables. Así pues, desde nuestro particular modo de ver las cosas, el problema que se le presenta a un investigador que dispone de múltiples medidas es decidir si es el diseño por sí el que determina su proceder; o si, por el contrario, condiciona su análisis al diseño más el modelo causal. En concreto, nosotros consideramos que esta última perspectiva puede ser de suma utilidad. Así las cosas, quizá sería conveniente antes de seguir adelante efectuar dos aclaraciones. En primer lugar, que no se trata de discutir, ni tampoco consideramos que sea pertinente hacerlo, si es mejor utilizar procedimientos analíticos propios del enfoque causal o la técnica del AMVAR a la hora de probar la hipótesis nula de igualdad multivariada, pues a nadie se le oculta que llevar a cabo un análisis factorial multivariado utilizando el LISREL es mu-

cho más complicado y laborioso (véase Vallejo, 1989) que lo que resulta aplicar un AMVAR, sin afectar, por supuesto, a las inferencias causales concernientes a las diferencias entre grupos. En segundo lugar, que tampoco nos vamos a referir al conjunto de técnicas que pueden ser utilizadas una vez que la hipótesis nula ha sido rechazada después de utilizar un AMVAR, proceder que consideramos propio de investigaciones de carácter más exploratorio.

Una vez llegados a este punto, y habiendo decidido investigar, no sólo la contribución única de cada variable, sino también la contribución común compartida por éstas, se nos presenta una doble cuestión. Por un lado, ¿cómo saber el número de factores necesarios para relacionar las variables independientes con las variables dependientes? y, por otro, ¿cómo determinar cuál de las diferentes estrategias disponibles para analizar medidas múltiples debemos utilizar?

Con el propósito de facilitar la comprensión de las preguntas planteadas, antes de pasar a responderlas, consideraremos la secuencia de modelos causales, que vamos a pasar a describir e ilustrar en los apartados que siguen.

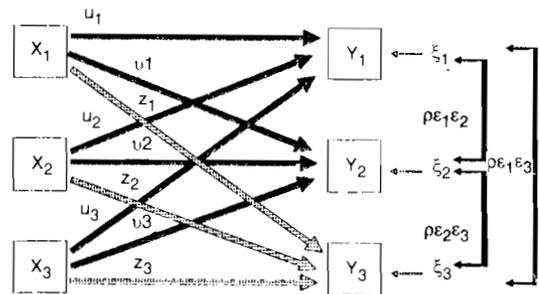


Figura 3. Diagrama causal para un diseño factorial con dos variables independientes y tres variables de medida.

El diagrama de la figura 3 ilustra un modelo causal para un diseño factorial 2×2 con tres variables dependientes. En este modelo las variables independientes cualitativas son definidas mediante una codificación de efectos. Así la variable X_1 representa al efecto principal de A y es definido como 1 si la observación aparece en a_1 y como -1 si una observación aparece en a_2 . La variable X_2 representa el efecto principal de B, y es definido de manera similar, con 1 si una observación aparece en b_1 y con -1 si aparece en b_2 . La variable interacción X_3 se obtiene multiplicando los valores de X_1 y X_2 para cada sujeto. Las tres variables dummy son mutuamente ortogonales; de ahí que no aparezca en el gráfico ninguna flecha que las interrelacione. Y_1 , Y_2 , e Y_3 representan las tres variables dependientes y ϵ_1 , ϵ_2 y ϵ_3 sus respectivos términos de error, éstos están unidos por flechas bidireccionales ya que se asume que están correlacionados. Los coeficientes

u_1, u_2 y u_3 ; v_1, v_2 y v_3 y z_1, z_2 y z_3 son las correlaciones de X_1, X_2 y X_3 con Y_1, Y_2 e Y_3 , respectivamente. En general en el caso de p grupos y q variables dependientes dispondremos de $p-1$ variables de codificación y de $q(p-1)$ coeficientes. La hipótesis nula multivariada sería declarada falsa si y sólo si al menos uno de los $q(p-1)$ coeficientes es distinto de cero en la población.

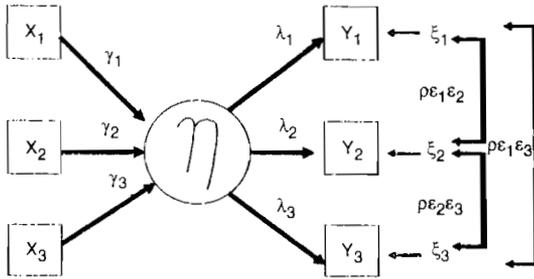


Figura 4a. Diagrama causal para un diseño factorial con dos variables independientes y tres criterios de medida. En el que se asume que los tratamientos causan una variable no observada de la cual las variables dependientes son sólo indicadores.

El modelo causal subyacente al diseño 2×2 con tres variables de medida representado en la figura 4a incluye un nuevo tipo de variable que hemos simbolizado por η , la cual no es observable. Por todo ello tenemos que este diagrama puede ser conceptualizado como un modelo de causas e indicadores múltiples con un solo constructo latente (MIMIC). En este tipo de modelo no hay término de error asociado con η , ya que se asume que el constructo es una función exacta de X_1, X_2 y X_3 . Aunque en este ejemplo hemos presupuesto que solamente un constructo media los efectos de las dos variables independientes en las tres dependientes, en la práctica puede ocurrir que modelos más complicados den cuenta de una manera más adecuada de nuestros

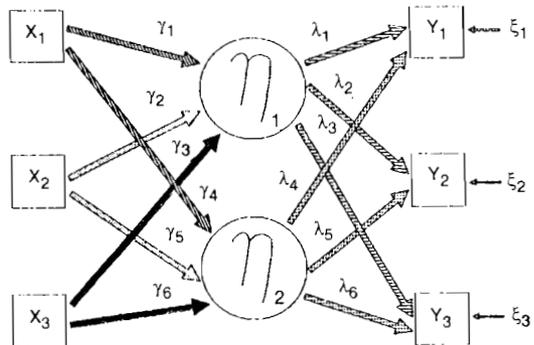


Figura 4b. Diagrama causal en el que un diseño factorial multivariado ha sido formulado como un modelo multicausa-multiefecto con dos variables no observadas.

datos. Hauser (1973), Stapleton (1978) y Hauser y Goldberger (1975) analizan, teórica y empíricamente, la extensión del modelo a situaciones como la esquematizada en la figura 4b.

Por último, el diagrama de la figura 5 ilustra un modelo causal para un diseño factorial de 2×2 con tres variables dependientes intercorrelacionadas. La peculiaridad de esta esquematización es que hemos asumido que cada variable de medida contribuye de manera similar a la variada compuesta; es decir, que hemos partido de que en la combinación lineal ponderada las cargas son aproximadamente iguales.

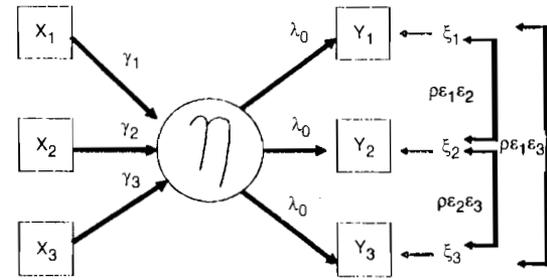


Figura 5. Diagrama causal de un diseño factorial multivariado en el que se asume que una variable no observada media la relación entre las observadas, pero son similares valores en las cargas canónicas cruzadas.

Los modelos expuestos pueden ser agrupados en dos bloques claramente diferenciados. El primero de ellos corresponde al gráfico de la figura 3; en el modelo esquematizado en este diagrama el investigador estaría interesado en averiguar si sus grupos difieren, mientras que en los modelos esquematizados en los gráficos de las figuras restantes al investigador también le interesaría saber cuánto y cómo difieren; en definitiva, aquí no sólo nos motivaría conocer si los tratamientos tienen o no efecto, sino que, en el caso de tenerlo, además de desear algo acerca de la forma de su diferencia también nos agradaría saber algo acerca de su magnitud. Es decir, que en los modelos del segundo bloque de diagramas el objetivo último del investigador no se limitaría a examinar la covariación entre los dos conjuntos de variables, sino que su trabajo estaría orientado a la identificación de las dimensiones relevantes que subyacen en el dominio de las p medidas multivariadas, de modo que el número de variables quede reducido al menor número posible de constructos significativos. Para ello, en lugar de evaluar si hay relación significativa entre los dos conjuntos de variables observadas, bien sea por la vía univariada, o bien por medio del AMVAR, tal y como lo desarrolló inicialmente Wilk (1932) en base al cálculo de una razón de verosimilitudes, se analizarían los factores o combinación lineal de variables de medida a través del análisis factorial o mediante el examen de las cargas canónicas obtenidas al realizar lo que Rao (1955) denominó análisis factorial canóni-

co. Particularmente, a nosotros este último procedimiento nos parece de especial interés, sobre todo, teniendo presente el contexto donde está insertado este trabajo.

A partir de lo dicho en el punto anterior, la determinación del número de dimensiones para explicar las diferencias entre los grupos se sigue de una manera casi inmediata del propio procedimiento señalado. De este modo, si se asume que no media ningún factor entre los dos conjuntos de variables, la mayor correlación canónica de la población no debe diferir significativamente de cero al probar esta hipótesis por medio del enfoque de la raíz característica mayor (Morrison, 1976). En el caso de que un factor sea suficiente para explicar la relación entre los dos conjuntos de variables la mayor correlación canónica sí diferirá significativamente de cero, mientras que el resto de los valores no lo harán, y así sucesivamente. En el supuesto de que los objetivos del experimentador se orientasen a obtener múltiples combinaciones lineales de variables ordenadas jerárquicamente y mutuamente incorrelacionadas, pero no necesariamente biortogonales, no sería necesario más que modificar ligeramente este enfoque, como se desprende del trabajo de Lambert, Wildt y Durand (1988).

Finalmente, por lo que respecta a la cuestión planteada cuál de las diferentes estrategias analíticas disponibles hemos de utilizar, nosotros vamos a tratar de responder circunscribiendo el problema a tres posibles situaciones, a saber:

1. Que las variables se presentan incorrelacionadas, o, aun estándolo, no sea la dimensionalidad subyacente de interés.
2. Que las variables se hallen correlacionadas y sea capital indagar en la estructura multidimensional de los criterios.
3. Que las variables de medida se encuentren interrelacionadas, pero que éstas contribuyan al constructo subyacente de forma similar.

El primer caso, que como puede observarse se corresponde con el gráfico de la figura 3, la forma más típica de efectuar el análisis es llevar a cabo un AVAR o un análisis de la regresión con cada una de las variables dependientes separadamente; sin embargo, lo más correcto sería analizar las múltiples variables dependientes simultáneamente mediante la ejecución de un AMVAR o de un análisis de la correlación canónica. Con todo, es importante recordar aquí algunas de las consideraciones que fueron hechas a la hora de hablar de las ventajas estadísticas obtenidas mediante el uso de pruebas multivariadas, y tener bien presente que la postura defendida por nosotros no fue la de llevar a cabo análisis multivariados por el simple hecho de tener múltiples variables. En concreto, apuntábamos que en el caso en que las variables estuvieran incorrelacionadas, bien porque por necesidades metodológicas (correlación y fiabilidad irregular) se las hubiera transformado, bien porque lo estaban originalmente, la utilización de un AMVAR no tenía ventajas espe-

ciales sobre la realización de pruebas univariadas separadas con el nivel α ajustado, inclusive estas últimas podrían protegernos contra un exceso de liberalidad en relación con el rechazo de alguna H_0 verdadera, si no se aceptaba la H_0 multivariada. Por otra parte, también señalábamos que cuando el tamaño de muestra y el número de tratamientos era reducido el AMVAR podía ser menos poderoso en términos estadísticos que las pruebas univariadas separadas; a propósito de lo cual, no hay que olvidar que las pruebas estadísticas univariadas, entendiéndose AVAR o regresión múltiple, no sólo son más sencillas de efectuar y, sobre todo, de interpretar, sino que también proporcionan una serie de cálculos auxiliares, ausentes en los paquetes estadísticos utilizados usualmente para ejecutar el AMVAR o la correlación canónica, como, por ejemplo, los errores estándar de los coeficientes de regresión o la contribución de cada variable a los cambios operados en el coeficiente de determinación.

En la segunda situación, esto es, la que se corresponde con los gráficos representados en las figuras 4a y 4b pueden ser utilizadas distintas técnicas analíticas. Una forma correcta de hacerlo sería analizar conjuntamente las múltiples variables dependientes por medio del AMVAR o de la correlación canónica, de modo que se capte lo especificado en el diagrama de la figura 4a, esto es, la explicación de que la correlación entre los dos conjuntos de variables observadas está mediada por una o más no observadas.

Otra forma de proceder analíticamente dentro de esta situación implicaría llevar a cabo un análisis univariado (AVAR, análisis de la regresión canónica) utilizando como variable(s) dependiente(s) el factor o factores definido(s) a través de alguna de las técnicas de reducción de datos o simplificación estructural, o bien mediante el análisis factorial canónico. Sobre todo, en aquellos casos, como el representado en la figura 4b, en los que no existan otras formas añadidas de covariación entre las variables, al margen, claro está, del propio factor. También cabe la posibilidad, como han puesto de manifiesto los trabajos citados anteriormente de Hauser y Goldberger (1971), Jöreskog y Golberger (1975) y de Stapleton (1978), de llevar a cabo simultáneamente la estimación del factor o factores y de probar el efecto de los tratamientos en dichas estimaciones. Mediante este procedimiento el problema anotado anteriormente de especificar la significación estadística de los coeficientes es superado. Pues, para muestras moderadas, se puede obtener la varianza del error estimada que cada parámetro tiene asociada y utilizar ésta para evaluar su significación.

Por último, en el caso de que las variables dependientes estén correlacionadas, como sucede con las del diagrama mostrado en la figura 5, pero la contribución de cada variable de medida (supuestamente registrada en la misma escala) a la combinación lineal ponderada, o variable compuesta, sea similar un investigador también puede contemplar la posibilidad de efectuar un AVAR de medidas repetidas, ya que en este tipo de análisis lo que se hace es esti-

mar el efecto de los tratamientos en el promedio de las diferentes variables de medida. Dado que este modelo de análisis pondera a todas las variables de medida por igual, necesitamos saber si las variables de nuestro modelo cumplen con este requisito; para ello es necesario que la hipótesis nula de ausencia de interacción tratamientos por variables criterios sea aceptada. Sin embargo, para que esta prueba esté libre de sesgos debe ser satisfecha en el delicado supuesto de que los sujetos no interactúan con los tratamientos (véase Vallejo, 1990 para una presentación más amplia de este problema.)

A modo de conclusión

El objetivo del presente trabajo ha consistido en discutir una serie de aspectos relacionados con el análisis de diseños experimentales con variables de respuesta múltiple. Para ello hemos expuesto, en primer lugar, cómo diferentes razones de tipo teórico, metodológico y estadístico nos capacitan para descubrir relaciones complejas entre las variables, para comprender mejor el fenómeno bajo investigación y para protegernos más correctamente frente a la posibilidad de cometer errores. En segundo lugar, hemos considerado una serie de criterios para evitar hacer uso indiscriminado de medidas múltiples. Y, por último, hemos debatido acerca de las técnicas analíticas a utilizar en el caso de disponer de un conjunto de criterios. Esta polémica no la hemos enfocado desde la contribución única de cada variable a la separación de los grupos, sino más bien teniendo presente la contribución conjunta; de ahí que para ello nos hayamos centrado, principalmente, en descubrir los factores que median la relación entre los dos conjuntos de variables observadas y en qué estrategia analítica utilizar en función de la forma de las intercorrelaciones encontradas.

En consecuencia, a la vista de lo dicho, podemos concluir diciendo que, mediante el uso de los diseños experimentales multivariados, se logran dos ventajas muy importantes. Por un lado, al controlarse la tasa de error Tipo I disminuyen las posibilidades de capitalizar los descubrimientos sobre el azar e incrementándose, por tanto, la fiabilidad de éstos y, por otro, al consagrar nuestro esfuerzo al descubrimiento de variables multidimensionales con significación sustantiva contribuimos muy positivamente a la descripción teórica de los fenómenos sometidos a investigación y a la validez de las inferencias realizadas.

Referencias

- Alvin, D. F., y Tessler, R. C. (1974): Causal models, unobserved variables, and experimental data, *American Journal of Sociology*, 80, 58-86.
- Bagozzi, R. P.; Fornell, C., y Larcker, D. F. (1981): Canonical correlation analysis as a special case of a structural relations model, *Multivariate Behavioral Research*, 16, 437-454.
- Bird, K. D., y Hadzi-Pavlovic, D. (1983): Simultaneous test procedures and the choice of a test statistic in MANOVA, *Psychological Bulletin*, 93, 167-178.
- Bock, R. D. (1975): *Multivariate Statistical Methods in Behavioral Research*, New York, McGraw-Hill.
- Bray, J. H., y Maxwell, S. E. (1982): Analyzing and interpreting significant MANOVAS, *Review Educational Research*, 52, 340-367.
- Bray, J. H., y Maxwell, S. E. (1985): *Multivariate Analysis of Variance*, Beverly Hill, CA: Sage.
- Cole, D. A.; Howard, G. S., y Maxwell, S. E. (1981): Effects of monoversus multiple operationalization in construct validation efforts, *Journal of Consulting and Clinical Psychology*, 49, 395.
- Cook, T. D., y Campbell, D.T. (1979): *Cuasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago, Rand McNally.
- Das Gupta, S., y Perlman, M. (1974): Power of the non-central F test: Effect of additional variables on Hotelling's T^2 -test, *Journal of the American Statistical Association*, 69, 174-180.
- Finn, J. D. (1974): *A General Model for Multivariate Analysis*, New York, Holt, Rinehart and Winston.
- Gabriel, K. R. (1969): Simultaneous test procedures: Some theory of multiple comparisons, *Annals of Mathematical Statistics*, 40, 224-250.
- Harris, R. J. (1985): *A Primer of Multivariate Statistics* (2nd ed.), Orlando, F. L., Academic Press.
- Hauser, R. M. (1973): Disaggregating a social-psychological model for educational attainment. En A. S. Goldberger y O. D. Duncan (eds.): *Structural Equation Models in the Social Sciences*, New York, Seminar.
- Hauser, R. M., y Goldberger, A. S. (1971): The treatment of unobservable variables in path analysis. En H. L. Costner (ed.), *Sociological Methodology* (1977), San Francisco, Jossey and Bass.
- Huberty, C. J., y Morris, J. D. (1989): Multivariate analysis versus multiple univariate analysis, *Psychological Bulletin*, 105, 302-308.
- Hummel, T. J., y Sligo, J. R. (1971): Empirical comparison of univariate and multivariate analysis of variance procedures, *Psychological Bulletin*, 75, 49-57.
- Jöreskog, K. G., y Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable, *Journal of the American Statistical Association*, 10, 631-639.
- Judd, Ch. M., y Kenny, D. A. (1981): *Estimating the Effects of Social Interventions*, New York, Cambridge University Press.
- Kenny, D. A. (1979): *Correlation and Causality*, New York, John Wiley.
- Lambert, Z. V.; Wildt, A. R., y Durand, R. M. (1988): Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring inter-set associations, *Psychological Bulletin*, 104, 282-289.
- Li, Ch. Ch. (1977): *Introducción a la Estadística Experimental*, Barcelona, Omega.
- Miller, R. G. (1981): *Simultaneous Statistical Inferences* (2nd ed.), New York, Springer-Verlag.
- Morrison, D. F. (1976): *Multivariate Statistical Methods* (2nd ed.), New York, McGraw-Hill.
- Muller, K. E., y Peterson, B. L. (1984): Practical methods for computing power in testing the multivariate general linear hypothesis, *Computational Statistics and Data Analysis*, 2, 143-158.
- O'Grady, K. E. (1986): Simultaneous test and confidence intervals, *Behavior Research Methods, Instruments and Computers*, 18, 325-326.

- Olson, C. L. (1974): Comparative robustness of six test in multivariate analysis of variance, *Journal of the American Statistical Association*, 69, 894-908.
- Palmer, D. J.; Drummond, F.; Tollison, P., y Zinkgraf, S. A. (1984): An attributional investigation of performance outcomes for learning disabled and normal achieving students, *Journal of Special Education*, 16 (4), 21-32.
- Pedhazur, E. (1982): *Multiple Regression in Behavioral Research* (2nd ed.), New York, Holt, Rinehart and Winston.
- Pruzek, R. M. (1971): Methods and problems in the analysis of multivariate data, *Review of Educational Research*, 41, 163-190.
- Rao, C. R. (1955): Estimation and tests of significance in factor analysis, *Psychometrika*, 20, 93-111.
- Stapleton, D. C. (1978): Analyzing political participation data with a MIMIC models. En K. F. Schuessler (ed.): *Sociological Methodology*, San Francisco, Jossey-Bass.
- Stevens, J. (1986): *Applied Multivariate Statistical for the Social Sciences*, New Jersey, Lawrence Erlbaum.
- Strahan, R. I. (1982): Multivariate analysis on the problem of Type I error, *Journal of Counseling Psychology*, 29, 175-179.
- Tatsuoka, M. M. (1971): *Multivariate Analysis: Techniques for Educational and Psychological Research*, New York, John-Wiley.
- Timm, N. H. (1975). *Multivariate Analysis: With Applications in Education and Psychology*, Belmont, CA: Brooks/Cole.
- Vallejo, G. (1990): Análisis univariado y multivariado de los diseños de medidas repetidas de una sola muestra y de muestras divididas, *Serie Avances Metodológicos en Psicología*, 2, Barcelona, PPU.
- Vallejo, G. (1989): Análisis de un diseño factorial con múltiples variables dependientes por medio de un modelo lineal de relaciones estructurales. Manuscrito no publicado, Departamento de Psicología, Universidad de Oviedo.
- Wilk, S. S. (1932): Certain generalizations in the analysis of variance, *Biometrika*, 24, 471-494.
- Wilkinson, L. (1975): Response variable hypothesis in the multivariate analysis of variance, *Psychological Bulletin*, 82, 40-412.
- Zinkgraf, S. A. (1983): Performing factorial multivariate analysis of variance using canonical correlation analysis. *Educational and Psychological Measurement*, 43, 63-68.