

## Diseño de medidas repetidas con dependencia serial en el error bajo la violación de la asunción de homogeneidad

Paula Fernández(\*) y Guillermo Vallejo

Universidad de Oviedo

**Resumen:** Para un diseño experimental aleatorio con varias medidas consecutivas a través del tiempo en cada una de las unidades experimentales examinamos dos procedimientos de análisis. Uno de ellos consiste en abordar los datos aplicando el modelo mixto del AVAR con la estructura del error modelada mediante procesos AR. De otro modo, también abordamos el problema desde una perspectiva más general haciendo uso del enfoque multivariado de medidas repetidas. Simulamos datos mediante procedimientos de Monte Carlo para investigar el efecto que el incumplimiento de las asunciones de independencia, esfericidad y homogeneidad tiene sobre el grado de "sesgidez" de los parámetros estimados, sobre la probabilidad empírica de cometer errores Tipo I y sobre la potencia de prueba para cada uno de los procedimientos.

**Palabras clave:** Medidas repetidas, errores con dependencia serial, procesos autorregresivos.

**Title:** Repeated measures design with error serial dependence under violation of homogeneity assumption.

**Abstract:** Two analytical procedures for a randomized experimental design with several consecutive measures taken through time in each one of the experimental units, are examined in this paper. One of them consists in facing data applying AVAR mixed model with modeled error structure by AR processes. On the other side, the problem is also faced from a more general perspective, using a repeated measures multivariate approach. Data were simulated by Monte Carlo procedures to investigate the effect that no satisfaction of independence, sphericity and homogeneity assumptions have on the bias degree of estimated parameters, on the empirical probability to make type I errors, and on the power of statistical test of the two procedures.

**Key words:** Repeated measures, errors with serial dependence, autorregresives processes.

### Introducción

En las últimas décadas, un tema de debate sugerente ha sido el desarrollo de una alternativa a los diseños Fisherianos de grupos propios de la investigación experimental clásica. En estas circunstancias, los diseños de medidas repetidas representan una alternativa viable a los diseños de corte transversal, y es que uno de los métodos de investigación más frecuentemente utilizados es tomar observaciones repetidas sobre los mismos sujetos, bien sea en diferentes puntos del tiempo, bien sea bajo diferentes condiciones de una variable determinada. Para el análisis de estos datos se

hace uso de alguna de las diferentes variedades de los diseños de medidas repetidas y que hoy se encuentran entre los más comúnmente utilizados en la investigación médica, social y psicológica actual (Barcowski y Robey, 1984; Keselman y Keselman, 1990; Keselman, Lix y Keselman, 1993; Keselman, Carriere y Lix, 1993; Edwards, 1991; Wallenstein y Fleiss, 1979; Vasey y Thayer, 1987).

Aunque este tipo de diseños no están exentos de dificultades producidas por los efectos de secuencia (efectos de orden y efectos residuales), es meritorio de ellos que acentúan la validez de conclusión estadística debido a una mayor precisión en la estimación de la variable de tratamiento, mejorando de esta manera la potencia de prueba; facilitan la generalización de los efectos del tratamiento debido a la particular presentación de los

(\*) **Dirección para correspondencia:** Paula Fernández García. Facultad de Psicología. C/Aniceto Sela, s/n. 33005 Oviedo (España)

© Copyright 1996: Servicio de Publicaciones de la Universidad de Murcia, Murcia (España). ISSN: 0212-9728. Artículo recibido: 8-7-96, aceptado: 23-12-96.

niveles del mismo, mejorando así la validez externa; reducen drásticamente el tamaño de la muestra y además, en alguna situación particular, permiten elaborar la posible tendencia.

De acuerdo con esto, en los años recientes, los investigadores han enfocado su atención en determinar cual de las estrategias analíticas resulta más apropiada para este tipo de diseños. Tradicionalmente han sido analizados mediante el análisis univariado de la varianza (AVAR) con valores esperados mixtos, en los cuales todos los factores, excepto el factor bloque o sujeto, son considerados fijos. El AVAR asume la independencia de las puntuaciones, que los errores aleatorios de la población sigan una distribución normal y que la matriz de covarianzas sea común para cada uno de los grupos de tratamiento. De todos modos, debido a la naturaleza correlacionada de las medidas repetidas, el AVAR aplicado a estos diseños requiere el cumplimiento de una asunción adicional llamada esfericidad o circularidad. La cual existe sí y sólo sí las varianzas correspondientes a las diferencias entre cualquier par de tratamientos son iguales. La mayoría de los investigadores coinciden en señalar que cumplir la condición de circularidad es difícil de lograr en muchas situaciones, sobre todo cuando trabajamos con datos psicofisiológicos (Wilson, 1967, 1974; Jennings y Wood, 1976; Keselman, Rogan, Mendoza y Brien, 1980) y con datos de investigación en educación y psicología (estudios longitudinales, de crecimiento y aprendizaje clásico) (Davidson, 1972; Wilson, 1975b; Greenwald, 1976). La consecuencia del incumplimiento de esta asunción es que la razón  $F$  del efecto intrasujeto se vuelve excesivamente liberal, se encuentra positivamente sesgada y por tanto el valor real de significación excederá el valor nominal establecido por el experimentador, incrementándose de este modo la posibilidad de capitalizar sobre el azar.

Para solventar este problema se han desarrollado una gran variedad de procedimientos: Mín  $\epsilon$ , corresponde al mínimo

valor de  $\epsilon$  propuesto por Greenhouse y Geisser (1959);  $\hat{\epsilon}$ , estimador desarrollado por Greenhouse y Geisser (1959);  $\tilde{\epsilon}$ , estimador desarrollado por Huynh y Feldt (1976) y toda una gran variedad de ellos que consideran conjuntamente a  $\hat{\epsilon}$  y  $\tilde{\epsilon}$  desarrollados por diversos autores (Maxwell y Arvey, 1982; Quintana y Maxwell, 1985, 1994). También en el caso de que existan matrices de covarianza heterogéneas o con tamaños de grupos arbitrarios están disponibles los procedimientos de aproximación general (GA) y aproximación general mejorada (IGA) desarrollados por Huynh (1978) y aproximación general mejorada corregida (CIGA) Algina (1994) y Algina y Oshima (1995). Todos ellos tienen en común que incrementan el valor crítico al estar multiplicados los grados de libertad, del efecto principal y de las interacciones, por algún valor de  $\epsilon$ .

Cole y Grizzle (1966) partiendo del hecho de que las observaciones obtenidas en los diseños de medidas repetidas están correlacionadas y, por tanto, son esencialmente de naturaleza multivariada, sugieren que el método adecuado para analizar estos diseños es el procedimiento multivariado. El enfoque multivariado comparte con el enfoque univariado todas las asunciones, excepto que permite a la matriz de varianza-covarianza tener cualquier estructura.

Y es que en muchas investigaciones psicológicas la dependencia serial entre las observaciones tomadas desde una misma unidad de análisis en distintos momentos o intervalos temporales puede llegar a ser de considerable importancia, ya que las relaciones entre las observaciones sólo son entendibles en un cierto contexto dinámico en el que, junto a variables referidas a un mismo momento del tiempo, existen otros efectos desplazados, efectos que se presentan con intensidad creciente cuanto menor es el lapso del tiempo sobre el cual queda definido el modelo. Como se ha puesto de manifiesto en distintas investigaciones (Vallejo, 1986), típicamente, las observaciones registradas desde un mismo

sujeto, además de estar gradual y positivamente correlacionadas, presentan una matriz de varianzas-covarianzas entre las medidas repetidas que tiene una estructura Toeplitz (las puntuaciones más próximas presentan una correlación más elevada). De este modo, este fenómeno es contrario a las asunciones que subyacen al modelo mixto del análisis de varianza (AVAR) tal y como fue formulado inicialmente por Scheffé en 1956 que debe asumir que los errores asociados con cada puntuación son independientes. Sin embargo, el cumplimiento de esta asunción, que implica correlación constante o que  $C^* \Sigma C^*$  fuese una matriz diagonal, es difícil de lograr en los diseños de medidas repetidas con las observaciones ordenadas sobre el tiempo.

Una forma simple de una estructura de series de tiempo para datos igualmente espaciados es el modelo autorregresivo de primer orden, AR (1). Los datos generados a partir de un proceso estacionario AR (1) presentan varianza constante a través de los  $t$  puntos del tiempo y las correlaciones decrecen gradualmente sobre el tiempo. En presencia de dependencia serial positiva el término de error es menor de lo que debiera cometándose más errores de Tipo I de los debidos (el cual es el más frecuente observado en las Ciencias Sociales y Comportamentales, Maddala, 1977). En presencia de dependencia serial negativa el término de error se incrementa y los tradicionales contrastes  $t$  y  $F$  cometen más errores de Tipo II.

Este es el patrón usual que presentan las respuestas de los sujetos cuando se registran ordenadamente sobre el tiempo por no utilizar el principio de aleatorización, bien por imposibilidad o bien por interés. Cuando ésto suceda, las inferencias causales que realizamos con el habitual modelo del AVAR de medidas repetidas deberán ser cuestionadas, pues, si bien es cierto que los parámetros estimados de los efectos principales e interacciones están insesgados, la varianza del error y los correspondientes errores estándar de las medidas están estimados incorrectamente,

sesgados en función de los parámetros autorregresivos y/o de las medias móviles. Consecuentemente, los intervalos confidenciales de las medidas observadas pueden sistemáticamente estar subestimados o sobreestimados (Vallejo 1989, pp. 145-147; Suen, Lee y Owen, 1990).

## Objetivos

A tenor de lo comentado anteriormente, nos parece de gran importancia estudiar el comportamiento del enfoque multivariado (AMVAR) de medidas repetidas frente al correspondiente univariado, pero con la estructura del error modelada mediante algún proceso que tenga en cuenta la posible correlación serial (AR 1), amén de los procedimientos univariados ajustados: mín  $\epsilon$  (mínimo valor de  $\epsilon$ ); estimador  $\hat{\epsilon}$  (UVGG) y estimador  $\tilde{\epsilon}$  (UVHF). Por ello, el propósito del presente trabajo, será mostrar los resultados en relación a la probabilidad de cometer errores Tipo I y en relación a la potencia de prueba de los test estadísticos para cada uno de los procedimientos de análisis referidos, bajo el incumplimiento de tres supuestos requeridos por el modelo de diseño experimental de medidas repetidas: homogeneidad, esfericidad de la matriz de varianza-covarianza y equicorrelación entre todos los pares de observaciones de una misma unidad experimental, todo ello bajo dos situaciones diferentes de normalidad: normalidad con puntuaciones restringidas y normalidad con puntuaciones sin restringir.

Por lo que respecta a la homogeneidad y para diseños balanceados (que es el caso de nuestra investigación), distintos investigadores (Huynh, 1978; Keselman y Keselman, 1990; Keselman et al., 1993; Rogan, Keselman y Mendoza, 1979; Keselman et al., 1993; Keselman, Keselman y Lix, 1995) han puesto de relieve que la falta de homogeneidad en las matrices de varianza-covarianza de los grupos no afecta a los métodos univariados, sin embargo los mismos autores desaconsejan el procedimiento multivariado como método de

elección por su excesivo error de Tipo I. Detallando los resultados de la última investigación realizada por Keselman et al. (1995) encontramos que con respecto al efecto principal  $T^2$  fue liberal siempre,  $\hat{\epsilon}$  y  $\hat{\epsilon}/T^2$  controlan la tasa de error, siendo mejor  $\hat{\epsilon}$ . Con respecto a la interacción,  $\hat{\epsilon}$ ,  $PB$  y  $\hat{\epsilon}/PB$  controlan  $\alpha$ ,  $\hat{\epsilon}/PB$  obtiene el menor error, después  $\hat{\epsilon}$  y después  $PB$ . Keselman y Keselman (1990) observan un leve incremento del error en el AMVAR. Los mismos autores señalados anteriormente coinciden en que la potencia no se encuentra afectada por la existencia de heterogeneidad. Algina y Oshima (1995) estudian el comportamiento de los procedimientos  $\tilde{\epsilon}$ ,  $IGA$  y  $CIGA$  y concluyen que  $\tilde{\epsilon}$  provee mejor error de Tipo I para el efecto principal que los otros nombrados. No obstante, nuestro interés se centra en observar el comportamiento de los métodos de análisis bajo las condiciones de normalidad especificadas cuando existe ausencia de homogeneidad (en diferentes grados) y de equicorrelación entre todos los pares de observaciones de una misma unidad experimental.

La ausencia de esfericidad ha sido un tema sometido a mucho estudio y debate desde la década de los 60. Nuestro propósito no es, por tanto, estudiarla aisladamente, sino en presencia de dos distribuciones normales diferentes y en presencia y ausencia de la homogeneidad de las matrices de dispersión. También observaremos cual de los dos estadísticos,  $\hat{\epsilon}$  o  $\tilde{\epsilon}$ , es más robusto en las situaciones antes señaladas.

En relación con la correlación serial, hemos señalado anteriormente, que la ausencia de independencia entre las puntuaciones tomadas desde una misma unidad de análisis cuando es

observada repetidamente a través del tiempo, es consustancial en algunas situaciones donde el diseño de medidas repetidas es utilizado. Por este motivo, consideramos de capital importancia tener presente algún planteamiento que modele la correlación serial de los datos, de modo que si esta resulta ser significativa pueda ser eliminada (de las respuestas efectuadas por los sujetos). Por esta razón, es un objetivo primordial para nosotros, investigar en presencia de autocorrelación el comportamiento del enfoque multivariado de medidas repetidas frente al procedimiento univariado con la estructura del error modelada mediante algún proceso autorregresivo (AR) y frente a los procedimientos univariados ajustados. También nos interesa observar el comportamiento de estos métodos en ausencia de autocorrelación y ausencia de esfericidad. Todo ello bajo las dos condiciones de normalidad aludidas, en ausencia de esfericidad y en presencia y ausencia de correlación y homogeneidad.

### Parámetros de medias poblacionales y potencias de prueba a priori

Para determinar cómo afecta a la aceptación o rechazo incorrecto de las hipótesis referidas al tratamiento intra y a la interacción para el modelo

$$y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + (\alpha\beta)_{jk} + (\pi\beta)_{ki(j)} + \epsilon_{ijk}$$

, en las Tablas A1 y A2 presentamos los vectores de medias  $\mu$  utilizados y las potencias de prueba teóricas para los dos modelos del diseño estudiado:

- a.- Para la condición de  $H_0$  para el tratamiento y la interacción. Modelo no aditivo
- b.- Para el modelo aditivo

**Tabla A1:** Valores numéricos de los vectores de parámetros de medias poblacionales y de potencias de prueba *a priori*. Matriz  $\Sigma I$  ( $\phi = .80$ ;  $\epsilon = .47$ )

Casos $\mu_j$ $q^1$ $q^2$ $q^3$ $q^4$ $q^5$ $q^6$ $q^7$ $q^8$	SC	R. Homg	Potencias de Prueba teóricas					
			S.Cor.Esf		Esf.Cor		Multiv	
			.01	.05	.01	.05	.01	.05
Valores de los vectores de medias de la población bajo $H_0$	$SC_A = 0$	1:1						
	$SC_{S(A)} = 24.555$	1:1.5						
	$SC_{S(A)} = 30079.8$	1:2						
A $\mu'_{1}$ 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5	$SC_{S(A)} = 28127.8$							
	$SC_B = 0$							
	$SC_{AB} = 0$							
$\mu'_{2}$ 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5	$SC_{B \times S(A)} = 4725$	1:1						
	$SC_{B \times S(A)} = 5857.4$	1:1.5						
	$SC_{B \times S(A)} = 9371.8$	1:2						
Valores de los vectores de medias de la población bajo $H_1$ . Modelo aditivo	$SC_A = 3136$							
	$SC_{S(A)} = 24.555$	1:1						
	$SC_{S(A)} = 30079.8$	1:1.5						
B $\mu'_{1}$ 1 2 3 4 5 6 7 8	$SC_{S(A)} = 48127.8$							
	$SC_B = 1344$	1:2						
	$SC_{AB} = 0$							
$\mu'_{1}$ 8 9 10 11 12 13 14 15	$SC_{B \times S(A)} = 4725$	1:1	.99	.99	.65	.85	.40	.55
	$SC_{B \times S(A)} = 5857.41$	1:1.5	.99	.99	.64	.84	.31	.52
	$SC_{B \times S(A)} = 9371.8$	1:2	.87	.95	.55	.76	.22	.31

**Tabla A2:** Valores numéricos de los vectores de parámetros de medias poblacionales y de potencias de prueba a priori Matriz  $\Sigma II$  ( $\phi = 0$ ;  $\varepsilon = .47$ )

Casos $\mu_j$ $q^1$ $q^2$ $q^3$ $q^4$ $q^5$ $q^6$ $q^7$ $q^8$	SC	R. Homg	Potencias de Prueba teóricas					
			S.Cor.Esf		Esf.Cor		Multiv	
			.01	.05	.01	.05	.01	.05
Valores de los vectores de medias de la población bajo $H_0$	$SC_A = 0$ $SC_{S(A)} = 24.555$ $SC_{S(A)} = 30079.8$ $SC_{S(A)} = 48127.8$	1:1 1:1.5 1:2						
A $\mu'_1$ 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5	$SC_B = 0$ $SC_{AB} = 0$ $SC_{BxS(A)} = 4725$ $SC_{BxS(A)} = 5904.1$ $SC_{BxS(A)} = 9446.6$	1:1 1:1.5 1:2						
$\mu'_2$ 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5								
Valores de los vectores de medias de la población bajo $H_1$ . Modelo aditivo	$SC_A = 3136$ $SC_{S(A)} = 24063.9$ $SC_{S(A)} = 30079.8$ $SC_{S(A)} = 48127.8$ $SC_B = 1344$	1:1 1:1.5 1:2						
B $\mu'_1$ 1 2 3 4 5 6 7 8	$SC_{AB} = 0$ $SC_{BxS(A)} = 4723.31$ $SC_{BxS(A)} = 5904.14$ $SC_{BxS(A)} = 9446.63$	1:1 1:1.5 1:2						
$\mu'_1$ 8 9 10 11 12 13 14 15								
		1:1 1:1.5 1:2	.99 .99 .99 .99 .87 .95	.78 .92 .63 .84 .53 .74	.40 .55 .31 .52 .22 .31			

Respecto al cálculo de la potencia de prueba multivariada, así como el cálculo de la potencia de prueba univariada sin corregir la ausencia de esfericidad (ambas *a priori*) existe mucha literatura al respecto y por ello no nos detenemos en su presentación; no obstante, sí nos parece de capital importancia señalar que para calcular la potencia de prueba a priori univariada corrigiendo la falta de esfericidad hemos seguido el trabajo de Muller y Barton (1989). Estos autores exponen que en el caso en que  $\varepsilon \neq I$ , como sucede en el caso que nos ocupa, la F univariada ya no se distribuye exactamente como F no central, sino de una manera aproximada, es decir,

$$F_u \approx F[v_1 \varepsilon, v_2 \varepsilon; v_1 F_B \varepsilon].$$

Los resultados de Muller y Barton permiten aproximar la potencia para las pruebas de Geisser-Greenhouse y Huynh-Feldt. Por

ejemplo, para la prueba de Geisser-Grenhouse lo dicho conduce a utilizar

$F_{GG} \approx F[v_1 E(\hat{\varepsilon}), v_2 E(\hat{\varepsilon}); v_1 F_B \varepsilon]$ . Para ello Muller y Barton (1989) aproximan el valor esperado de  $\hat{\varepsilon}$  siguiendo el trabajo de Fujikoshi (1978).

### Procedimiento seguido en el cálculo de los datos y en la corrección de la correlación

En cada uno de los experimentos (que más adelante se detallarán) analizamos los datos correspondientes a los modelos aditivo y no aditivo con cada uno de los procedimientos analíticos como sigue:

- Para el análisis multivariado se ha hecho uso del paquete estadístico SPSS PC 5.0. De los cuatro estadísticos disponibles que este paquete dispensa para recoger el efecto del

tratamiento y de la interacción utilizamos la traza de Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) basándonos en los resultados de Olson (1974), Rogan et al. (1979) y Bird y Hadzi-Pavlovic (1983) que señalaban que es la más robusta de las pruebas y suficientemente sensible para detectar diferencias en poblaciones con alguna estructura de no centralidad. A pesar de nuestra inclinación por la prueba antes señalada sabemos del estudio de Stevens (1979) que encontró que la traza de Pillai-Bartlett, la  $\Delta$  de Wilks y la traza de Hotelling-Lawley son esencialmente igual de robustas. De este paquete estadístico también se han recogido los datos de los análisis univariados ajustados  $\hat{\epsilon}$ ;  $\tilde{\epsilon}$  y mín  $\epsilon$ .

- Para el análisis univariado de la varianza corrigiendo la correlación están disponibles varios programas (Azzalini, 1984; Jones, 1985 b; Andersen, Jensen y Schoy, 1989), con todo, se ha hecho uso del programa estadístico realizado por el profesor Richard H. Jones (1985 b) para los diseños de medidas parcialmente repetidas. Atendiendo al consejo de Azzalini y Frigo (1991) de la necesidad de tomar precaución a la hora de estimar la correlación serial, la elección ha estado subordinada a que este programa utiliza el método de máxima verosimilitud para este fin, y que según los estudios realizados, es el procedimiento que mejor se comporta en la mayoría de las situaciones.

Estos procedimientos de análisis que acabamos de detallar se hicieron bajo cada una de estas dos condiciones:

- a) Homoscedasticidad más correlación en la estructura del error
- b) Heteroscedasticidad y correlación arbitraria

De cada una de las dos condiciones efectuamos comparaciones en las siguientes áreas:

- 1.- Tasas de error de Tipo I
- 2.- Potencia de prueba
- 3.- Precisión de las estimaciones efectuadas

Medidas empíricas de la probabilidad de cometer errores Tipo I se obtuvieron tabulando el número de veces que cada estadístico excede su valor crítico cuando las diferencias en el

vector de medias es nula y dividiendo por el número de pruebas efectuadas. La potencia de prueba se derivó registrando el número de veces que la hipótesis nula es debidamente rechazada al nivel  $\alpha$  especificado y siempre en un modelo aditivo. Para estimar el grado de sesgo presente en los efectos del diseño se comparó el valor medio de los parámetros estimados, si el estimador es insesgado, entonces el valor promedio se aproximará estrechamente al verdadero valor del parámetro poblacional.

Conclusiones relativas a la eficacia y robustez de los diferentes procedimientos de estimación se obtuvieron comparando los respectivos errores estándar y aplicando el criterio liberal de robustez de Bradley (1978).

Cómputo de los errores estándar: los errores estándar reportados para estudiar las estimaciones empíricas de  $\alpha$  son calculados desde  $\sqrt{pq/m}$ , donde p es la probabilidad teórica del error de Tipo I, q es el complementario de p y m es el número de experimentos efectuados. Aquellas estimaciones empíricas de  $\alpha$  que estén fuera del intervalo  $\pm 2$  desviaciones estándar serán consideradas significativas; si este intervalo se supera por el lado superior las estimaciones serán llamadas liberales, si la estimación es inferior al mínimo valor del intervalo serán consideradas conservadoras.

Cómputo del criterio de robustez de Bradley (1978): de acuerdo con este criterio, y para que una prueba pueda ser considerada robusta, su proporción empírica del error de Tipo I ( $\hat{\alpha}$ ) debe estar contenida en el intervalo  $0.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$ .

## Metodología General

En orden a evaluar los objetivos expuestos, las propiedades de un modelo mixto del AVAR (con errores correlacionados) y las propiedades de un AMVAR de medidas repetidas serán investigadas por medio de datos simulados bajo diferentes situaciones. Las hipótesis a comparar son las referidas a un

diseño de medidas parcialmente repetidas (2x8) con ocho vectores de observaciones en cada uno de los dos diferentes grupos, participando 16 sujetos en cada uno de ellos. Para alcanzar este objetivo, desde distribuciones multivariadas normales extrajimos múltiples conjuntos de vectores pseudoaleatorios  $y'_{ij} [y_{ij1}, y_{ij2}, \dots, y_{ijk}]$  con un vector de medias  $\mu_j [\mu_{j1}, \mu_{j2}, \dots, \mu_{jk}]$  y matriz de varianzas-covarianzas  $\Sigma$

### Generación de datos

a) *Distribución Normal; puntuaciones restringidas*: El procedimiento seguido para la obtención de los vectores pseudoaleatorios  $y'_{ij}$  se efectuó en tres fases. En la primera, haciendo uso del método congruencial multiplicativo descrito por Naylor, Balintfy, Burdick y Chu (1966) generamos vectores de variables aleatorias independientes y uniformemente distribuidas entre cero y uno despreciando las puntuaciones que no se encontraban entre  $\pm 3$  desviaciones estándar. A continuación, haciendo uso del método de Teichrow descrito por Knuth (1969) las variables uniformes fueron convertidas en vectores de variables normales  $Z'_{ij} [Z_{ij1}, Z_{ij2}, \dots, Z_{ijk}]$ . Finalmente, los vectores de variables  $Z'_{ij}$  fueron transformados en vectores de observaciones  $y'_{ij}$  (donde  $y \sim N(\mu, \Sigma)$ ) siguiendo el método de Schauer y Stoller (1966) por medio de la ecuación  $y'_{ij} = Tz_{ij} + \mu$ , donde  $T$  es una matriz no única y es la descomposición triangular de  $\Sigma_j$ , que satisface la igualdad  $\Sigma = LL'$  (a menudo llamada factorización de Cholesky o método de la raíz cuadrada (Harman, 1967)). La precisión del procedimiento de normalización comprobada mediante los criterios de sesgo y curtosis aportados por el paquete SPSS PC 5.0 fue enteramente satisfactoria.

b) *Distribución normal; puntuaciones no restringidas*: el procedimiento seguido en la

obtención de los vectores pseudoaleatorios  $y'_{ij}$  se efectuó en tres fases, las mismas que las descritas en el apartado (a) con la diferencia de que en la primera fase se generaron vectores de variables aleatorias independientes y normalmente distribuidas pero no se desechan las puntuaciones que se encuentran fuera del intervalo  $\pm 3$  desviaciones estándar. De nuevo, la precisión del procedimiento de normalización comprobada mediante los criterios de sesgo y curtosis aportados por el paquete SPSS-PC 5.0 fue enteramente satisfactoria.

### Matrices de varianza-covarianza utilizadas

Las matrices que llamamos  $\Sigma I$  siguen un proceso Markov o autorregresivo de orden uno (en adelante se las señalará como matrices del modelo  $\Sigma I$ ) en concreto la matriz  $\Sigma I$  para una situación de homogeneidad (razón 1:1) ha sido construida desde

$$\Sigma = \sigma_\varepsilon^2 \left( \Omega - \frac{1}{q} j \right) + (\sigma_\varepsilon^2 + q \sigma_\pi^2) \frac{1}{q} j$$

$$(\phi = 0.8; \sigma_\varepsilon^2 = 18; \sigma_v^2 = 72)$$

la matriz  $\Sigma I$  para una heterogeneidad de razón 1:1.5

$$\Sigma = \sigma_\varepsilon^2 \left( \Omega - \frac{1}{q} j \right) + (\sigma_\varepsilon^2 + q \sigma_\pi^2) \frac{1}{q} j$$

$$(\phi = 0.8; \sigma_\varepsilon^2 = 27; \sigma_v^2 = 108)$$

y la matriz  $\Sigma I$  para una heterogeneidad de razón 1:2

$$\Sigma = \sigma_\varepsilon^2 \left( \Omega - \frac{1}{q} j \right) + (\sigma_\varepsilon^2 + q \sigma_\pi^2) \frac{1}{q} j$$

$$(\phi = 0.8; \sigma_\varepsilon^2 = 36; \sigma_v^2 = 144)$$

Aquí se asume que el proceso AR es estacionario. Como consecuencia inmediata de la estacionariedad, la varianza es constante para todos los valores de  $t$  y la covarianza entre dos cualesquiera términos de error depende únicamente de la distancia temporal entre ellos y no del momento del tiempo a que están referidos.

Las matrices de varianza-covarianza ( $\Sigma$ ) que denominaremos como matrices del

modelo ( $\Sigma_{II}$ ) además de presentar heterocedasticidad sobre el tiempo, los valores situados en las diagonales secundarias también difieren sensiblemente, es decir, tienen una correlación arbitraria.

Tanto las matrices a las que les subyace un modelo AR (I) ( $\Sigma_I$ ) como las matrices con correlación arbitraria ( $\Sigma_{II}$ ) comparten una misma desviación de la esfericidad  $\varepsilon = .46$ .

Los datos generados a partir de estas matrices nos van a permitir comparar las dos técnicas analíticas aludidas, así como el comportamiento del AVAR de medidas repetidas con la estructura del error modelada mediante algún proceso AR cuando la propiedad de igualdad de las diferentes diagonales de la matriz  $\Sigma$  es parcial o totalmente incumplida.

Estas matrices ( $\Sigma_I$  y  $\Sigma_{II}$ ) nos van a permitir también estudiar el supuesto de homogeneidad de las matrices de varianza-covarianza de los grupos del factor entresujetos. Este supuesto fue estudiado creando matrices para los niveles entresujetos cuyos elementos estaban en diferentes razones unos de otros. Esto es, cuando el supuesto de homogeneidad es satisfecho, los elementos de la matriz de un grupo fueron iguales a los elementos correspondientes a la otra matriz del otro grupo, será la condición 1:1. Dos condiciones en las cuales los elementos correspondientes a las matrices de varianza-covarianza de los grupos no fueron iguales unos de otros fueron elegidas para investigar la desviación del supuesto de homogeneidad: en la condición denominada de leve desviación de la homogeneidad los elementos de las matrices que conforman el factor entresujetos siguen una razón 1:1.5; en la condición denominada de gran desviación de la homogeneidad la razón es de 1:2.

Apuntar tan sólo lo siguiente: dado que los datos han sido simulados desde poblaciones conocidas y que cada conjunto de observaciones será analizado mediante los procedimientos (AR1, UVGG, UVHF, mín  $\varepsilon$  y AMVAR), contamos con una situación óptima para determinar el grado de robustez de

estos enfoques. De este modo, si un procedimiento es robusto con respecto a sus asunciones, los resultados esperados se obtendrán aunque éstas no se hayan cumplido. Por el contrario, si los supuestos son críticos y el enfoque no tolera desviarse de ellos los resultados diferirán de las conclusiones esperadas.

## Experimento 1

### Método

En este experimento de simulación Monte Carlo los datos siguen una distribución normal con puntuaciones restringidas entre  $\pm 3$  desviaciones estándar, y se han llevado a cabo 2000 análisis para cada uno de los dos modelos de un diseño factorial mixto con una variable entresujetos y una intrasujeto:

1.- Condición de  $H_0$  para el tratamiento y la interacción. Modelo no aditivo

2.- Modelo aditivo

Estos 2000 análisis se han ejecutado para cada una de las dos diferentes construcciones de las matrices de covarianza que conforman los datos:

1.- Homoscedasticidad mas correlación en la estructura del error (matrices  $\Sigma_I$ ). Siguen un proceso autorregresivo de orden uno ( $\phi = .8$ ,  $\varepsilon = .47$ ).

2.- Heteroscedasticidad y correlación arbitraria (matrices  $\Sigma_{II}$ ). ( $\phi = .0$ ,  $\varepsilon = .47$ ).

También hemos manipulado la homogeneidad de éstas en los dos grupos que forma la variable entresujetos para tres situaciones diferentes: homoscedasticidad (razón 1:1), heteroscedasticidad leve (razón 1:1.5) y heteroscedasticidad alta (razón 1:2).

La generación de los datos, la construcción de las matrices de varianza-covarianza así como los test estadísticos mediante los cuales han sido analizados los dos modelos del diseño, ya han sido expuestos en el apartado anterior.

**Resultados**

*Estimación empírica del error de Tipo I para el tratamiento. Modelo no aditivo. Tabla 1*

*Modelo ΣI:* Para un nivel de significación  $\alpha = .01$ , los procedimientos AR1 (procedimiento univariado con la correlación corregida) y UVGG mantienen el error en su nivel nominal, el AMVAR también, aunque con una tasa de error empírica muy inferior al nivel teórico ( $\hat{\alpha} = .006 ; \alpha = .01$ ); el procedimiento UVHF es liberal para el criterio del error estándar pero no para el criterio de robustez de Bradley. Para el nivel de significación  $\alpha = .05$  todos los procedimientos mantienen el error en su nivel nominal y para  $\alpha = .10$  también, excepto

UVGG que es conservador. El procedimiento univariado ajustado mediante el mínimo valor de  $\epsilon$ , es conservador para todos los niveles de significación utilizados.

Si las matrices de dispersión presentan una heterogeneidad leve todos los procedimientos mantienen el error en su nivel nominal para  $\alpha = .01$ , salvo UVHF que es liberal. Para  $\alpha = .05$  todos los procedimientos mantienen el error en su nivel nominal y para  $\alpha = .10$  también, excepto UVGG y UVHF que son conservadores. Si la dispersión es alta todos los procedimientos se ajustan bien en los tres niveles de significación utilizados. El procedimiento mín  $\epsilon$  en ausencia de homogeneidad se manifiesta siempre conservador.

**Tabla 1:** Estimación empírica de la tasa de error Tipo I para el tratamiento. Modelo no aditivo. ( $\Sigma I: \phi = .80, \epsilon = .47$ ) ( $\Sigma II: \phi = 0, \epsilon = .47$ ) N = 2000 análisis

		$\alpha=.01$			$\alpha=.05$			$\alpha=.10$											
		1:1	1:1.5	1:2	1:1	1:1.5	1:2	1:1	1:1.5	1:2									
$\Sigma I$	AR (1)	.013	.015	.008	.019	.010	.014	.048	.053	.049	.054	.053	.056	.092	.102	.092	.100	.096	.100
	$\hat{\epsilon}$	.012	.013	.011	.013	.011	.012	.041	.045	.039	.043	.047	.05	.075+	.082+	.074+	.080+	.081	.086
	$\tilde{\epsilon}$	.014*	.016*	.016*	.017*	.014	.016*	.045	.05	.031	.051	.053	.057	.082	.094	.077+	.091	.089	.094
	mín $\epsilon$							.009+	.010+	.009+	.01+	.009+	.01+	.034+	.038+	.023+	.034+	.036+	.037+
	AMVAR	.006	.007	.007	.008	.007	.008	.042	.046	.045	.049	.053	.057	.086	.094	.088	.095	.104	.11
	ES	.0022248			.0048733			.006708											
	Bradley	.005-.015			.025-.075			.05-.15											
$\Sigma II$	AR (1)	.035*	.039*	.037*	.039*	.425*	.044*	.096*	.105*	.092*	.01*	.097*	.101*	.132*	.145*	.133*	.143*	.149*	.155*
	$\hat{\epsilon}$	.007	.008	.007	.007	.009	.01	.040	.043	.042	.042	.046	.048	.085+	.092	.092	.088	.084+	.087
	$\tilde{\epsilon}$	.009	.010	.012	.012	.011	.011	.051	.055	.049	.052	.053	.055	.09	.104	.092	.099	.102	.105
	mín $\epsilon$					.0005+		.007+	.007+	.006+	.006+	.009+	.008+	.031+	.034+	.031+	.033+	.093	.039+
	AMVAR	.009	.010	.009	.010	.013	.014	.036+	.039	.043	.045	.048	.051	.080+	.087	.088	.094	.103	.108
	ES	.0022248			.0048733			.006708											
	Bradley	.005-.015			.025-.075			.05-.15											

Comportamiento liberal;+ Comportamiento conservador.

De igual manera para todas las tablas:

Puntuaciones en negrilla: Experimento 2, normalidad con puntuaciones sin restringir;

Puntuaciones normales: Experimento 1, normalidad con puntuaciones restringidas.

*Modelo ΣII:* Para un  $\alpha = .01$  los procedimientos UVGG, UVHF y AMVAR mantienen la tasa del error controlada por debajo de su nivel nominal. Para el  $\alpha = .05$  y  $.10$  los procedimientos UVGG y UVHF se comportan igual que para el  $\alpha = .01$ , sin embargo, el AMVAR es conservador. El procedimiento mín  $\epsilon$  y AR1 son conservador y

liberal respectivamente para todos los niveles de significación utilizados.

En ausencia de homogeneidad todos los procedimientos mantienen el error en su nivel nominal excepto mín  $\epsilon$  que es conservador y AR1 que es liberal para todos los niveles de significación utilizados.

Estimación empírica del error de Tipo I para la interacción. Modelo no aditivo. Tabla 2

Modelo  $\Sigma I$ : Todos los procedimientos mantienen el error en su nivel nominal al  $\alpha = .01$  y  $\alpha = .05$ . Para  $\alpha = .10$  el procedimiento UVGG es conservador. El procedimiento mín  $\epsilon$  es conservador para todos los niveles de significación utilizados.

En situación de heterogeneidad todos los procedimientos, excepto mín  $\epsilon$  que es conservador, mantienen el error en su nivel nominal.

Modelo  $\Sigma II$ : El procedimiento mín  $\epsilon$  es siempre conservador, y el procedimiento AR1

es siempre liberal para los tres niveles de significación utilizados. Los demás procedimientos se ajustan al nivel nominal si éste es .01; para  $\alpha = .05$  y .10 también, excepto UVGG que es conservador.

En ausencia de homogeneidad todos los procedimientos univariados mantienen el error en su nivel nominal para todos los niveles de significación utilizados, excepto mín  $\epsilon$  que siempre es conservador y AR1 que siempre es liberal. El AMVAR se manifiesta liberal sólo para  $\alpha = .10$ .

Tabla 2: Estimación empírica de la tasa de error Tipo I para la interacción. Modelo no aditivo ( $\Sigma I$ :  $\phi = .80$ ,  $\epsilon = .47$ ) ( $\Sigma II$ :  $\phi = 0$ ,  $\epsilon = .47$ ) N=2000 análisis

		$\alpha=.01$						$\alpha=.05$						$\alpha=.10$					
		1:1		1:1.5		1:2		1:1		1:1.5		1:2		1:1		1:1.5		1:2	
$\Sigma I$	AR (1)	.014	.015*	.012	.013	.009	.01	.051	.055	.016	.054	.056	.058	.102	.111	.065	.107	.103	.108
	$\hat{\epsilon}$	.011	.012	.009	.009	.011	.012	.041	.045	.046	.048	.046	.049	.078+	.087	.086	.09	.087	.092
	$\tilde{\epsilon}$	.014	.015*	.013	.014	.014	.015	.048	.053	.051	.056	.057	.06	.088	.096	.088	.098	.095	.094
	mín $\epsilon$	.0005+		.0005+		.0005+		.008+	.01+	.008+	.008+	.009+	.009+	.033+	.037+	.035+	.037+	.034+	.035+
	AMVAR	.011	.012	.013	.014	.013	.011	.043	.0465	.043	.047	.049	.054	.081+	.098	.086	.098	.092	.099
	ES	.0022248						.0048733						.006708					
	Bradley	.005-.015						.025-.075						.05-.15					
$\Sigma II$	AR (1)	.032*	.036*	.033*	.035*	.033*	.035*	.090*	.089*	.088*	.094*	.096*	.101*	.141*	.153*	.139*	.148*	.153*	.160*
	$\hat{\epsilon}$	.007	.008	.008	.009	.009	.009	.036	.038+	.035	.036	.038	.040	.073+	.084+	.085+	.088	.089	.093
	$\tilde{\epsilon}$	.011	.013	.011	.012	.012	.013	.040	.043	.044	.046	.045	.046	.088	.096	.092	.096	.106	.11
	mín $\epsilon$	.0005+						.005+	.005+	.004+	.005+	.006+	.006+	.029+	.031+	.027+	.028+	.033+	.034+
	AMVAR	.010	.011	.011	.011	.0469	.012	.043	.052	.052	.053	.092*	.053	.100	.101	.101	.105	.14*	.107
	ES	.0022248						.0048733						.006708					
	Bradley	.005-.015						.025-.075						.05-.15					

Estimación empírica del error de Tipo I para la interacción. Modelo aditivo. Tabla 3

**Tabla 3:** Estimación empírica de la tasa de error Tipo I para la interacción. Modelo aditivo. ( $\Sigma I: \phi = .80, \epsilon = .47$ ) ( $\Sigma II: \phi = 0, \epsilon = .47$ ) N=2000 análisis

		$\alpha=.01$			$\alpha=.05$			$\alpha=.10$											
		1:1	1:1.5	1:2	1:1	1:1.5	1:2	1:1	1:1.5	1:2									
$\Sigma I$	AR (1)	.013	<b>.013</b>	.011	<b>.013</b>	.009	<b>.01</b>	.052	<b>.056</b>	.050	<b>.054</b>	.056	<b>.058</b>	.105	<b>.111</b>	.101	<b>.107</b>	.104	<b>.108</b>
	$\hat{\epsilon}$	.012	<b>.012</b>	.008	<b>.009</b>	.011	<b>.013</b>	.044	<b>.045</b>	.045	<b>.048</b>	.047	<b>.048</b>	.084	<b>.087</b>	.084	<b>.09</b>	.089	<b>.092</b>
	$\tilde{\epsilon}$	.014	<b>.015*</b>	.013	<b>.014</b>	.014	<b>.015*</b>	.051	<b>.053</b>	.053	<b>.056</b>	.057	<b>.06</b>	.092	<b>.09</b>	.093	<b>.098</b>	.095	<b>.099</b>
	min $\epsilon$		<b>.0005+</b>		<b>.0005+</b>		<b>.001+</b>	.009+	<b>.01+</b>	.007+	<b>.008+</b>	.009+	<b>.009+</b>	.035+	<b>.098</b>	.035+	<b>.095</b>	.033+	<b>.096</b>
	AMVAR	.012	<b>.012</b>	.013	<b>.014</b>	.014	<b>.01</b>	.044	<b>.046</b>	.044	<b>.047</b>	.051	<b>.052</b>	.084	<b>.098</b>	.088	<b>.095</b>	.092	<b>.096</b>
	ES	.0022248			.0048733			.006708											
	Bradley	.005-.015			.025-.075			.05-.15											
$\Sigma II$	AR (1)	.064*	<b>.036*</b>	.035*	<b>.037*</b>	.034*	<b>.035*</b>	.090*	<b>.098*</b>	.091*	<b>.097*</b>	.094*	<b>.098*</b>	.141*	<b>.153*</b>	.145*	<b>.153*</b>	.147*	<b>.155*</b>
	$\hat{\epsilon}$	.006	<b>.007</b>	.008	<b>.009</b>	.009	<b>.009</b>	.034+	<b>.037+</b>	.036+	<b>.038+</b>	.035+	<b>.037+</b>	.076+	<b>.08+</b>	.082+	<b>.08+</b>	.081+	<b>.085+</b>
	$\tilde{\epsilon}$	.010	<b>.011</b>	.011	<b>.012</b>	.010	<b>.011*</b>	.039	<b>.042</b>	.045	<b>.048</b>	.042	<b>.043</b>	.088	<b>.096</b>	.092	<b>.095</b>	.093	<b>.098</b>
	min $\epsilon$							.004+	<b>.004+</b>	.004+	<b>.005+</b>	.005+	<b>.006+</b>	.028+	<b>.031+</b>	.028+	<b>.03+</b>	.029+	<b>.032</b>
	AMVAR	.009	<b>.010</b>	.011	<b>.012</b>	.038*	<b>.012</b>	.046	<b>.049</b>	.047	<b>.049</b>	.076*	<b>.052</b>	.094	<b>.099</b>	.105	<b>.11</b>	.128*	<b>.106</b>
	ES	.0022248			.0048733			.006708											
	Bradley	.005-.015			.025-.075			.05-.15											

*Modelo  $\Sigma I$ :* Todos los procedimientos mantienen el error en su nivel nominal para todos los valores de  $\alpha$  en presencia de homogeneidad y de heterogeneidad, excepto el procedimiento mín  $\epsilon$ , que es siempre conservador.

*Modelo  $\Sigma II$ :* Los procedimientos AR1 y mín  $\epsilon$  son siempre liberal y conservador respectivamente. Para  $\alpha = .01$  los demás procedimientos se ajustan a su valor nominal, para  $\alpha = .05$  y  $.10$  también, excepto UVGG que es conservador.

Si existe heterogeneidad leve todos los procedimientos se ajustan a su nivel nominal, salvo UVGG que es conservador. Si la heterogeneidad es alta todos los procedimientos son robustos excepto el AMVAR.

*Estimación empírica de la potencia de prueba. Modelo aditivo. Tabla 4*

*Modelo  $\Sigma I$ :* Los procedimientos univariados ajustados y AMVAR tienen una potencia empírica superior a la potencia teórica

( $1 - \hat{\beta} = .908, 1 - \beta = .65$ ) y ( $1 - \hat{\beta} = .44, 1 - \beta = .40$ ) para UVGG y AMVAR respectivamente al  $\alpha = .01$ ) para todos los niveles de significación utilizados, pero son los primeros los que son más poderosos para detectar los efectos del tratamiento. El procedimiento AR1 ( $1 - \hat{\beta} = .63, 1 - \beta = .99; \alpha = .01$ ), siendo más poderoso que el AMVAR, tiene una potencia inferior al nivel teórico.

La heterogeneidad de los grupos afecta a la potencia reduciéndola para la razón 1:1.5 pero se ve considerablemente mejorada si la razón es 1:2.

*Modelo  $\Sigma II$ :* Todos los procedimientos tienen una potencia superior al nivel teórico e igualmente elevada.

De nuevo respecto a la heterogeneidad de los grupos, si ésta es de una razón 1:1.5 reduce un poco la potencia ( $1 - \hat{\beta} = 1$  para la razón 1:1, frente a  $1 - \hat{\beta} = .974$  para la razón 1:1.5 en UVGG), pero ésta se vuelve a recuperar si la razón es 1:2.

**Tabla 4:** Estimación empírica de la potencia de prueba para el tratamiento. Modelo aditivo.  
( $\Sigma I$ :  $\phi = .80, \epsilon = .47$ ) ( $\Sigma II$ :  $\phi = 0, \epsilon = .47$ ) N=2000 análisis

		$\alpha=.01$			$\alpha=.05$			$\alpha=.10$											
		1:1	1:1.5	1:2	1:1	1:1.5	1:2	1:1	1:1.5	1:2									
$\Sigma I$	AR (1)	.638	<b>.604</b>	.612	<b>.564</b>	.844	<b>.826</b>	.855	<b>.857</b>	.839	<b>.771</b>	.961	<b>.946</b>	.949	<b>.897</b>	.939	<b>.863</b>	.999	<b>.978</b>
	$\hat{\mathcal{E}}$	.908	<b>.868</b>	.889	<b>.816</b>	.994	<b>.954</b>	1	<b>.962</b>	1	<b>.935</b>	1	<b>.977</b>	.983	<b>.968</b>			.975	
	$\tilde{\mathcal{E}}$	.927	<b>.893</b>	.900	<b>.846</b>	.999	<b>.961</b>	.999	<b>.966</b>	1	<b>.944</b>	1	<b>.979</b>	.984	<b>.972</b>			.980	
	min $\epsilon$	.561	<b>.529</b>	.447	<b>.422</b>	.801	<b>.795</b>	.922	<b>.876</b>	.859	<b>.893</b>	.912	<b>.955</b>	.999	<b>.96</b>	.980	<b>.929</b>	.932	<b>.975</b>
	AMVAR	.448	<b>.425</b>	.394	<b>.371</b>	.669	<b>.649</b>	.752	<b>.71</b>	.685	<b>.638</b>	.897	<b>.861</b>	.883	<b>.834</b>	.834	<b>.778</b>	.969	<b>.905</b>
	ES	.0022248			.0048733			.006708											
	Bradley	.005-.015			.025-.075			.05-.15											
$\Sigma II$	AR (1)	1	1	1	<b>.999</b>	1	1	1											
	$\hat{\mathcal{E}}$	1	<b>.997</b>	.974	<b>.975</b>	1	1	1	<b>.999</b>	1									
	$\tilde{\mathcal{E}}$	1	<b>.9998</b>	1	<b>.987</b>	1	1	1	1	1									
	min $\epsilon$	.782	<b>.742</b>	.794	<b>.781</b>	.925	<b>.956</b>	1	<b>.998</b>	1	<b>.972</b>	1	1	1	<b>.999</b>				
	AMVAR	1	1	1	1	1	1												
	ES	.0022248			.0048733			.006708											
	Bradley	.005-.015			.025-.075			.05-.15											

Los parámetros  $\epsilon$  y  $\phi$  se ajustan casi perfectamente con un valor medio de  $\epsilon \approx .45993$  y  $\phi \approx .819$  en el conjunto de los 2000 análisis efectuados para este experimento.

## Experimento 2

### Método

En este experimento de simulación Monte Carlo los datos siguen una distribución normal con puntuaciones sin restringir y se han llevado a cabo 2000 análisis para cada uno de los dos modelos de un diseño factorial mixto con una variable entresujetos y una intrasujeto:

1.- Condición de  $H_0$  para el tratamiento y la interacción. Modelo no aditivo

2.- Modelo aditivo

Estos 2000 análisis se han ejecutado para cada una de las dos diferentes construcciones de las matrices de covarianza que conforman los datos:

1.- Homoscedasticidad más correlación en la estructura del error (matrices  $\Sigma I$ ). Siguen un proceso autorregresivo de orden uno ( $\phi = .8, \epsilon = .47$ ).

2.- Heteroscedasticidad y correlación arbitraria (matrices  $\Sigma II$ ). ( $\phi = 0, \epsilon = .47$ ).

También hemos manipulado la homogeneidad de éstas en los dos grupos que forma la variable entresujetos para tres situaciones diferentes: homoscedasticidad (razón 1:1), heteroscedasticidad leve (razón 1:1.5) y heteroscedasticidad alta (razón 1:2).

La generación de los datos, la construcción de las matrices de varianza-covarianza así como los test estadísticos mediante los cuales han sido analizados los dos modelos del diseño, ya han sido expuestos en el apartado anterior.

*Estimación empírica del error de Tipo I para el tratamiento. Modelo no aditivo. Tabla 1*

*Modelo  $\Sigma I$ :* Para un  $\alpha = .01$  los procedimientos AR1 y UVHF tienen un comportamiento liberal; UVGG y AMVAR

mantienen el error en su nivel nominal siendo este último más conservador. Para un  $\alpha = .05$  todos los procedimientos mantienen el error ajustado y para  $\alpha = .10$  también, excepto UVGG que es conservador. El procedimiento mín  $\mathcal{E}$  es siempre conservador.

En presencia de heterogeneidad leve estos resultados no varían. Si la heterogeneidad es alta, todos los procedimientos excepto mín  $\mathcal{E}$  que es conservador, ajustan el error empírico al error prefijado.

*Modelo  $\Sigma II$* : Los procedimientos UVGG, UVHF y AMVAR mantienen el error en su nivel nominal para todos los niveles de significación empleados. El procedimiento mín  $\mathcal{E}$  es siempre conservador y el AR1 siempre liberal.

Estos resultados no varían para los dos grados de heterogeneidad utilizados.

*Estimación empírica del error de Tipo I para la interacción. Modelo no aditivo. Tabla 2*

*Modelo  $\Sigma I$* : Todos los procedimientos mantienen el error en su nivel nominal para  $\alpha = .05$  y  $.10$ ; si  $\alpha = .01$ , AR1 y UVHF tienen un comportamiento liberal. El procedimiento mín  $\mathcal{E}$  siempre es conservador.

En presencia de heterogeneidad todos los procedimientos mantienen el error en su nivel nominal a excepción de UVHF, que en presencia de heterogeneidad alta y para  $\alpha = .01$  manifiesta un comportamiento liberal.

*Modelo  $\Sigma II$* : Los procedimientos AR1 y mín  $\mathcal{E}$  son siempre liberal y conservador respectivamente, los demás procedimientos mantienen el error en su nivel nominal para todos los niveles de significación, excepto UVGG que es conservador para  $\alpha = .05$  y  $.10$ .

En presencia de heterogeneidad se mantienen estos resultados.

*Estimación empírica del error de Tipo I para la interacción. Modelo aditivo. Tabla 3*

Tanto para el modelo  $\Sigma I$  como para el modelo  $\Sigma II$  los resultados son los mismos que para la interacción en el modelo no aditivo.

*Estimación empírica de la potencia de prueba. Modelo aditivo. Tabla 4*

*Modelo  $\Sigma I$* : Los procedimientos univariados ajustados y AMVAR tienen una potencia empírica superior a la potencia teórica para todos los niveles de significación utilizados ( $1 - \hat{\beta} = .86, 1 - \beta = .65$  y  $1 - \hat{\beta} = .42, 1 - \beta = .40$  para UVGG y AMVAR respectivamente para  $\alpha = .01$ ), pero son los primeros más poderosos para detectar los efectos del tratamiento. El procedimiento AR1, si bien es más poderoso que el AMVAR, tiene una potencia inferior al nivel teórico ( $1 - \hat{\beta} = .60, 1 - \beta = .99; \alpha = .01$ ).

La heterogeneidad de los grupos afecta a la potencia reduciéndola para la razón 1:1.5, pero se ve considerablemente mejorada si la razón es 1:2.

*Modelo  $\Sigma III$* : Todos los procedimientos tienen una potencia superior al nivel teórico e igualmente elevada.

De nuevo la heterogeneidad de los grupos, si ésta es una razón 1:1.5, reduce un poco la potencia ( $1 - \hat{\beta} = .997$  para la razón 1:1, frente a  $1 - \hat{\beta} = .975$  para la razón 1:1.5 en UVGG), pero ésta se vuelve a recuperar si la razón es 1:2.

## Discusión

Recordando los objetivos que nos planteamos y a la luz de los resultados obtenidos, podemos señalar los siguientes puntos:

**En presencia de correlación serial elevada y desviación de esfericidad severa**

( $\phi = .80$  y  $\varepsilon = .47$ )

*Error de Tipo I para el tratamiento:*

- El AVAR con dependencia serial corregida (AR 1) ( $\hat{\alpha} = .0139, \alpha = .01$ ) sólo se ajusta mejor que el AMVAR ( $\hat{\alpha} = .006, \alpha = .01$ ) en situación de

normalidad perfecta (puntuaciones restringidas) para todos los niveles de significación, y en este caso, el AR1 ( $\hat{\alpha} = .0139; \alpha = .01$ ) no es mejor que el procedimiento UVGG ( $\hat{\alpha} = .0139; \alpha = .01$ ).

- En presencia de normalidad con puntuaciones sin restringir, más acercado a la realidad, sólo UVGG ( $\hat{\alpha} = .013$ ) y AMVAR ( $\hat{\alpha} = .007$ ) ajustan el error empírico al teórico para  $\alpha = .01$ , y para  $\alpha = .05$  y  $.10$  son el AMVAR y UVHF los que poseen el mejor comportamiento.

- El procedimiento mín  $\mathcal{E}$  es conservador en la mayoría de las situaciones apoyando los resultados obtenidos en situación de independencia por Collier, Baker, Mandeville y Hayes (1967), Rogan et al. (1979), Keselman y Keselman (1990), Maxwell y Arvey (1982).

- Dado que existe ausencia de esfericidad severa ( $\epsilon = .47$ ) el procedimiento UVGG se ajusta mejor que el procedimiento UVHF y el AMVAR sólo para  $\alpha = .01$ , siendo estos últimos procedimientos liberal y conservador respectivamente para este nivel de significación. Para  $\alpha = .05$  y  $.10$ , UVGG, UVHF y AMVAR mantienen el error en su nivel nominal, pero se ajustan mejor los dos últimos, pecando UVGG un poco de conservador. Dado que  $q=8$  estamos de acuerdo con las afirmaciones realizadas por Keselman y Keselman (1990), Maxwell y Arvey (1982), Rogan et al. (1979), Keselman, Keselman y Lix (1995). Barcikowski y Robey (1984) y Looney y Stanley (1989) (todos estos estudios en situación de independencia) que muestran que no hay una clara ventaja entre  $\hat{\mathcal{E}}$  y la prueba multivariada. Acudiendo a la investigación realizada por Collier y otros (1967), encontramos que su estudio revela que si  $\epsilon \leq .75$ ,  $\hat{\mathcal{E}}$  ofrece una estimación menos sesgada que  $\tilde{\mathcal{E}}$ , pero si contamos con 15 observaciones por grupo, estas estimaciones estaban menos sesgadas y las diferencias entre  $\hat{\mathcal{E}}$  y  $\tilde{\mathcal{E}}$  se acortaban. Teniendo en cuenta que nosotros contamos con 16 observaciones por grupo y que hemos observado aún una distancia entre estos dos procedimientos, sería

bueno en una próxima investigación y atendiendo a la reciente polémica despertada por Lecoutre (1991), comprobar si estas diferencias entre UVGG y UVHF siguen existiendo si sustituimos UVHF por su corrección en esta situación de dependencia, al igual que afirman Algina y Oshima (1995), que en su investigación (situación de independencia) encuentran que  $\tilde{\mathcal{E}}$  experimenta el mejor comportamiento. O si por el contrario confirmamos los hallazgos de Quintana y Maxwell (1994) al respecto y la diferencia entre UVGG y UVHF se presenta menos palpable.

- La ausencia de homogeneidad entre las matrices de dispersión de los grupos de tratamiento actúa de una forma extraña, de igual manera para los dos tipos de distribución estudiados; para los procedimientos univariados éste se mantiene igual que cuando existe homogeneidad o incluso disminuye si la heterogeneidad es media; si es alta, el error siempre aumenta, aunque nunca sobrepasa los criterios de robustez, lo cual nos hace pensar que la dependencia serial afecta a los procedimientos univariados cuando se incumple el supuesto de homogeneidad. Para el AMVAR, conforme incrementa la heterogeneidad incrementa siempre el error, a pesar de ello, mantiene el error en su nivel nominal, lo cual nos hace pensar que en situación de dependencia serial este procedimiento es robusto frente a la ausencia de homogeneidad.

#### *Error de Tipo I para la interacción:*

- Es indistinto utilizar cualquier procedimiento, ya que no le afecta significativamente la forma de distribución de la normalidad manteniéndose todos los procedimientos dentro del criterio de robustez de Bradley, no obstante, es el AMVAR el que experimenta el mejor comportamiento.

- La ausencia de homogeneidad entre las matrices de dispersión de los grupos de tratamiento actúa de una forma extraña; para los procedimientos univariados éste se

mantiene igual que cuando existe homogeneidad o incluso disminuye si la heterogeneidad es media; si es alta, el error siempre aumenta. Para el AMVAR, conforme incrementa la heterogeneidad incrementa siempre el error. A pesar de ello, todos los procedimientos mantienen el error en su nivel nominal. Con lo cual haríamos las mismas observaciones que para el tratamiento cuando se incumple este supuesto.

- En situación de dependencia, como denominador común y teniendo en cuenta los resultados de esta investigación, es aconsejable utilizar el AMVAR.

*Potencia:*

- Todos los procedimientos tienen una potencia superior a su valor teórico y son los procedimientos univariados (UVHF>UVGG) más poderosos que el AMVAR.

- La ausencia de homogeneidad afecta a la potencia de un modo curioso. La empeora si la razón de heterogeneidad es 1:1.5 (aun manteniéndose superior siempre a su valor teórico) y la favorece si la razón es 1:2.

**En presencia de correlación arbitraria y ausencia de esfericidad severa**  
( $\phi = 0.0$  y  $\varepsilon = .47$ )

*Error de Tipo I para el tratamiento:*

- Los procedimientos univariados ajustados y AMVAR se acomodan igual de bien al error nominal, en este orden de mejor a peor, UVHF, AMVAR y UVGG. Este resultado es de esperar, ya que los procedimientos UVGG y UVHF corrigen la falta de esfericidad en una situación donde no hay correlación y por tanto no se ven perjudicados, y el AMVAR está operando en una situación óptima para él y no le afecta la ausencia de esfericidad. Estos mismos resultados son encontrados por Keselman et al. (1995); Keselman et al. (1993), Keselman y Keselman (1990) y Maxwell y Arvey (1982).

- El procedimiento AR1 es excesivamente liberal, pero éste es tan sólo un artefacto del procedimiento de análisis, pues al tener la matriz correlación cero y falta de esfericidad, el modelo multivariado tiene en cuenta este hecho, mientras que el AR1 se comporta como el usual modelo mixto del AVAR sin corregir los grados de libertad por la falta de esfericidad.

- La ausencia de homogeneidad entre las matrices de dispersión de los grupos de tratamiento actúa de una forma extraña de igual manera para los dos tipos de distribución estudiados; para los procedimientos univariados éste se mantiene igual que cuando existe homogeneidad o incluso disminuye si la heterogeneidad es media; si es alta, el error siempre aumenta aunque nunca sobrepasa los criterios de robustez. Sin embargo, en las investigaciones realizadas por Huynh, 1978; Keselman y Keselman, 1990; Keselman et al., 1993; Rogan et al., 1979; Keselman et al., 1993; Keselman et al., 1995, se desprendía que estos procedimientos no se veían afectados, lo cual nos empuja a replicar y a hacer otros experimentos donde la ausencia de homogeneidad sea más acusada. Para el AMVAR, conforme incrementa la heterogeneidad incrementa siempre el error, a pesar de ello, lo mantiene en su nivel nominal. De esta forma desmentimos las afirmaciones llevadas a cabo por Huynh, 1978; Keselman y Keselman, 1990; Keselman et al., 1993; Rogan et al., 1979; Keselman et al., 1993; Keselman et al., 1995, de las cuales se desprende que este procedimiento no es robusto frente a la ausencia de homogeneidad. No obstante, harían falta otros estudios para comprobar si los niveles de heterogeneidad empleados por nosotros son o no muy leves.

- AR1 y mín  $\mathcal{E}$  son siempre liberal y conservador respectivamente.

*Error de Tipo I para la interacción:*

- Es aconsejable utilizar cualquier procedimiento con excepción de AR1 que

siempre es liberal y mín  $\mathcal{E}$  que siempre es conservador.

- Como denominador común y teniendo en cuenta los resultados de esta investigación, es aconsejable utilizar el AMVAR o UVHF. Compartimos los resultados encontrados en otros estudios puntualizando que Keselman y Keselman (1990) aconsejan utilizar UVHF y Keselman et al. (1993) y Keselman et al. (1995) la estrategia combinada  $\hat{\epsilon} / T^2$ , después  $\hat{\mathcal{E}}$  y después  $T^2$ .

- En ausencia de homogeneidad, los distintos procedimientos se comportan de la misma manera que para la interacción en situación de dependencia serial. Estos resultados concuerdan con los hallados por Huynh, 1978; Keselman y Keselman, 1990; Keselman et al., 1993; Rogan et al., 1979; Keselman et al., 1993; Keselman et al., 1995.

#### Potencia:

- Todos los procedimientos tienen la potencia igualmente elevada (siempre es mayor para el AMVAR) y no hay un procedimiento significativamente mejor que otro, lo mismo que encontraron O'Brien y Kaiser (1985), Stevens (1986), Barcikowski y Robey (1974), Rogan et al. (1979); Maxwell y Delaney (1990), Keselman y et al. (1993).

- La ausencia de homogeneidad afecta a la potencia de un modo curioso. La empeora si la razón de heterogeneidad es 1:1.5 (aun manteniéndose siempre superior a su valor teórico) y la favorece si la razón es 1:2. Con lo cual no estamos de acuerdo con Keselman et al. (1995) que encuentran una potencia no afectada cuando se incumple el supuesto de homogeneidad.

## Conclusiones

### Teniendo en cuenta la forma de distribución

- Cuando pasamos de puntuaciones restringidas a sin restringir aumenta un poco el error aunque casi nunca es significativo, a

excepción del AR1 y UVHF que son liberales para  $\alpha = .01$  en  $\Sigma I$ . En  $\Sigma II$  el error también experimenta un incremento, pero en menor medida que para  $\Sigma I$ .

- El procedimiento multivariado, de igual manera para los dos tipos de matrices utilizadas, experimenta un incremento progreso del error conforme incrementa la heterogeneidad, aunque siempre permanece en el intervalo de  $\pm 2$  SE. En curioso que en la situación  $\Sigma II$ , este procedimiento siempre tiene un error inferior (salvo para  $\alpha = .01$  que puede ser igual) que los demás procedimientos estudiados, a excepción de mín  $\mathcal{E}$ , y el incremento de la heterogeneidad favorece a este procedimiento, ya que hace que se acerque más al nivel nominal que nos proponemos, ajustándose a éste perfectamente cuando la heterogeneidad es alta.

- Con respecto a la interacción en los dos tipos de distribución estudiados, en situación  $\Sigma I$  los procedimientos se comportan de la misma forma que lo hacen para el tratamiento, sin embargo para  $\Sigma II$  el AMVAR siempre manifiesta un error igual o mayor que los procedimientos univariados ajustados, siendo liberal para  $\alpha = .10$  si las puntuaciones son restringidas.

- Cuando pasamos de puntuaciones restringidas a sin restringir, la potencia experimenta un pequeño decremento en los dos tipos de matrices utilizados para todos los procedimientos, excepto para AR1 y AMVAR.

- La potencia disminuye para la razón de heterogeneidad 1:1.5 y se recupera para la razón 1:2.

- El procedimiento UVGG siempre es más conservador que UVHF tanto para el tratamiento como para la interacción en los dos tipos de distribución estudiados, y conforme incrementa el nivel de significación, mayor es la proporción en que se aleja de éste, ocasionando, a veces, que si para el nivel nominal .01 ó .05 mantiene el error en su nivel nominal, para .10 sea conservador.

- Todos los procedimientos se encuentran afectados en presencia de heterogeneidad, sin embargo, para los procedimientos univariados

no hay ninguna regla clara, en cambio, el AMVAR sí experimenta un incremento progresivo del error conforme la heterogeneidad aumenta, aunque rara vez hace variar el resultado que obtiene en situación de homogeneidad.

### **Teniendo en cuenta la dependencia o independencia de las puntuaciones**

*En cuanto al error de Tipo I para el tratamiento cuando existe homogeneidad:* En ausencia de esfericidad severa ( $\epsilon = .47$ ) y normalidad con puntuaciones sin restringir (la misma normalidad utilizada en los estudios publicados) UVGG ( $\hat{\alpha} = .13$ ) sólomente se ajusta mejor que los demás procedimientos para  $\alpha = .01$ , lo mismo que es situación de no autocorrelación. Sin embargo el uso de UVHF o AMVAR en situación de dependencia para  $\alpha = .01$  hace que estas pruebas sean liberal y conservadora respectivamente para este nivel de significación, situación en la cual si no hubiese dependencia controlarían el error de Tipo I aunque peor que UVGG. Para  $\alpha = .05$  y  $\alpha = .10$  todos los procedimientos ajustan el error empírico al error teórico tanto en situación de dependencia como de independencia.

*En cuanto al error de Tipo I para la interacción cuando existe homogeneidad:* Tanto en presencia como en ausencia de autocorrelación todos los procedimientos aquí estudiados tienen un buen comportamiento.

*En cuanto a la potencia cuando existe homogeneidad:* En situación de dependencia, aunque todos los procedimientos tienen una potencia superior a su valor teórico, son más poderosos los procedimientos univariados (de mejor a peor: UVHF, UVGG y AMVAR), al contrario de lo que se esperaría. Cuando no

existe dependencia serial, aunque todos los procedimientos tienen una potencia muy alta, es el AMVAR el más poderoso.

*Error de Tipo I en presencia de heterogeneidad:* lo mismo para la interacción que para el tratamiento, la dependencia serial parece que sí influye en los procedimientos univariados cuando se incumple el supuesto de homogeneidad, aunque en nuestra investigación no ha sido de una forma significativa. Sería necesario estudiar otros niveles de autocorrelación junto con otras situaciones de heterogeneidad para hacer conclusiones más exactas. El AMVAR es robusto frente a la ausencia de homogeneidad cuando existe dependencia. Es curioso que los mismos resultados son obtenidos cuando no hay dependencia, lo cual, de nuevo, nos hace pensar en realizar nuevas investigaciones.

*Potencia en presencia de heterogeneidad:* Si existe dependencia serial son más potentes los procedimientos univariados y cuando existe independencia, es el procedimiento multivariado el más potente. Lo que también parece es que ésta oscila, tanto a la alta como a la baja, para diferentes niveles de heterogeneidad. De nuevo, esto nos invita a seguir realizando nuevas investigaciones.

Por último señalar que los resultados presentados aquí sólo podrán ser interpretados y generalizados dentro de las limitaciones de este estudio, marcados por el valor de los parámetros utilizados ( $\mu$ ,  $\Sigma$ ,  $\alpha$ ,  $\beta$ ,  $\phi$  y  $\epsilon$ ), tamaño de la muestra, así como por la forma de distribución subyacente. El camino para refutar o confirmar estos resultados es llevar a cabo más estudios de naturaleza similar con otros diseños de medidas repetidas, otras estructuras de  $\Sigma$  y variando los tamaños muestrales.

## Referencias bibliográficas

- Algina, J. (1994). Some alternative approximate tests for a Split Plot design. *Multivariate Behavioral Research*, 29 (4), 365-384.
- Algina, J. y Oshima, T.C. (1995). An improved general approximation test for the main effect in a Split-Plot design. *British Journal of Mathematical and Statistical Psychology*, 48, 149-160.
- Andersen, A.H., Jensen, E. y Schou, G. (1981). Two-way analysis of variance with correlated errors. *International Statistical Review*, 49, 153-167.
- Azzalini, A. (1984). Estimation and hypothesis testing for collections of autoregressive time series. *Biometrika*, 71 (1), 85-90.
- Azzalini, A. y Frigo, A.C. (1991). An explicit nearly unbiased estimate of the AR(1) parameter for repeated measurements. *Journal Time Series Analysis*, 12, 273-281.
- Barcikowski, R.S. y Robey, R.R. (1984). Decisions in single group repeated measures analysis: Statistical test and tree computer packages. *The American Statistician* 38, 148-151.
- Bartlett, M.S. (1939). A note on a test of significance in multivariate analysis. *Proceedings of the Cambridge Philosophical Society*, 35, 180-185.
- 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.
- Bradley, J.V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.
- Cole, V.W.L. y Grizzle, J.E. (1966). Applications of multivariate analysis of variance to repeated measures experiments. *Biometrics*, 41, 505-514.
- Collier, R.O., Baker, F.B., Mandeville, G.K. y Hayes, T.F. (1967). Estimates of tests size for several tests procedures based on conventional variance ratios in the repeated measures design. *Psychometrika*, 32, 339-353.
- Davidson, M.L. (1972). Univariate versus multivariate test in repeated measures experiments. *Psychological Bulletin*, 77, 446-452.
- Edwards, L.K. (1991). Fitting a Serial Correlation Pattern to Repeated Observations. *Journal of Educational Statistics*, 16, 53-73.
- Fujikoshi, Y. (1978). Asymptotic Expansions for the Distributions of Some Functions of the Latent Roots of matrices in three situations. *Journal of Multivariate Analysis*, 8, 63-72.
- Greenhouse, S.W. y Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-102.
- Greenwald, A.G. (1976). Within-subjects designs: To use or not to use?. *Psychological Bulletin*, 83 (2), 314-320.
- Harman, L.G. (1967). An generalized simplex for factor analysis. *Psychometrika*, 20, 173-192.
- Huynh, H. (1978). Some approximate test for repeated measurement designs. *Psychometrika*, 43, 161-175.
- Huynh, H. y Feldt, L.S. (1976). Estimation of the Box correction for degrees of freedom from sample data in the randomized block and Split-plot designs. *Journal of Educational Statistics*, 1, 1582-1589.
- Jennings, J.R. y Wood, C.C. (1979). The epsilon adjusted procedure for repeated measures analyses of variance. *Psychophysiology*, 13, 277-278.
- Jones, R.H. (1985 b). Repeated measures, interventions, and time series analysis. *Psychoneuroendocrinology*, 10 (1), 5-15.
- Keselman, J.C. y Keselman, H.J. (1990). Analysing unbalanced repeated measures designs. *British Journal of Mathematical and Statistical Psychology*, 43, 265-282.
- Keselman, J.C., Lix, L.M. y Keselman, H.J. (1993). *The analysis of repeated measurements: A quantitative research synthesis*. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta.
- Keselman, H.J., Carriere y Lix (1993). Testing repeated measures hypotheses when covariance matrices are heterogeneous. *Journal of Educational Statistics*, Vol 18, 4, 305-319.
- Keselman, H.J., Rogan, J.C., Mendoza, J.L. y Brien, L.J. (1980). Testing the validity conditions of repeated measures F test. *Psychological Bulletin*, 87, 479-481.
- Keselman, H.J., Keselman, J.C. y Lix L.M. (1995). The analysis of repeated measurements: Univariate tests, multivariate tests, or both?. *British Journal of Mathematical and Statistical Psychology*, 48, 319-338.
- Kirk, R.E. (1968). *Experimental design*. Monterey. Ca.: Brooks/Cole.
- Knuth, D.E. (1969). *Seminumerical Algorithms: The Art of Computer Programming*. Massachusetts: Addison-Wesley.
- Lecoutre, B. (1991). A correction for the  $\tilde{E}$  approximate test in repeated measures designs with two or more independent groups designs with two or more independent groups. *Journal of Educational Statistics*, 16, 371-372.
- Looney, S.W. y Stanley, W.B. (1989). Exploratory repeated measures analysis for two or more groups: Review and update. *American Statistics*, 43, 200-225.
- Maddala, G.S. (1977). *Econometrics*. New York: McGraw-Hill. (Traducido de la edición en inglés por la propia editorial).
- Maxwell, S.E. y Arvey, R.D. (1982). Small sample profile analysis with many variables. *Psychological Bulletin*, 92, 778-785.
- Muller, K.E. y Barton, C.N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association*, 84, 549-556.
- Naylor, H.T., Balintfy, J.L., Burdick, D.S. y Chu, K. (1966). *Computer Simulation Techniques*. New York: John Wiley.

- Neyman, J. y Scott, E.L. (1959). Estochastic models of population dynamics. *Science*, 130, 303-308.
- O'Brien, K.G. y Kaiser, M.K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, 97, 316-333.
- Olson, C.L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 69, 894-908.
- Pillai, K.C.S. (1955). Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, 26, 117-121.
- Quintana, S.M. y Maxwell, S.E. (1985). A better than average estimate of  $\mathcal{E}$ . Paper presented at the annual meeting of the American Educational Research Association.
- Quintana, S.M. y Maxwell, S.E. (1994). A Monte Carlo comparison of seven  $\mathcal{E}$ -Adjustment procedures in repeated measures designs with small sample sizes. *Journal of Educational Statistics*, 19 (1), 57-71.
- Rogan, J.C., Keselman, H.J. y Mendoza, J.L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, 32, 269-286.
- Schauer, E.M. y Stoller, S.D. (1966). On the generation of normal random vectors. *Technometrics*, 4, 279-290.
- Scheffé, H. (1959). *The analysis of variance*. John Wiley and Sons, Inc., New York.
- Stevens, J.P. (1986). *Applied Multivariate Statistics for the Social Sciences*. New Jersey, LEA.
- Suen, H.K., Lee, P.S y Owen, S.V. (1990). Effects of autocorrelation on Single-Subject Single-facet crossed-design generalizability assessment. *Behavioural Assessment*, 12, 305-315.
- Vallejo, G. (1986). Aplicación del análisis de series temporales en diseños con N=1: Consideraciones generales. *Revista Española de Terapia del Comportamiento*, 4, 1-29.
- Vallejo, G. (1986). Procedimientos simplificados de análisis en los diseños de series temporales interrumpidas. (modelos estáticos). *Revista Española de Terapia del Comportamiento*, 4, 114-148.
- Vallejo, G. (1989). Regresión de series de tiempo con mediciones igualmente espaciadas. *Anuario de Psicología*, 43, 126-155.
- Vasey, M.W. y Thayer, J.F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A Multivariate Solution. *Psychophysiology*, 24 (4), 479-486.
- Wilson, R.S. (1967). Analysis of automatic reaction patterns. *Psychophysiology*, 4, 125-142.
- Wilson, R.S. (1974). CARDIVAR: The statistical analysis of heart rate data. *Psychophysiology*, 11, 76-85.
- Wilson, R.S. (1975 b). Analysis of developmental data: Comparison among alternative methods. *Developmental Psychology*, 11 (6), 676-680.
- Wallenstein, S y Fleiss, J.L. (1979). Repeated measurements analysis of variance when the correlations have a certain pattern. *Psychometrika*, 42 (2), 229-233.