QÜESTIIÓ, vol. 22, 3, p. 443-468, 1998

ALGUNAS SOLUCIONES APROXIMADAS PARA DISEÑOS SPLIT-PLOT CON MATRICES DE COVARIANZA ARBITRARIAS

G. VALLEJO SECO
J.R. ESCUDERO GARCÍA
Universidad de Oviedo*

El presente trabajo revisa con cierto detalle diversos tipos de análisis para diseños split-plot que carecen del mismo número de unidades experimentales dentro de cada grupo y en los que se incumple con el supuesto de esfericidad multimuestral. Específicamente, adaptando el enfoque multivariado de aproximar los grados de libertad desarrollado por Johansen (1980) y el procedimiento de la aproximación general mejorada corregida basado en Huynh (1980) se muestra cómo obtener análisis robustos y poderosos a la hora de probar los efectos principales y la interacción, así como hipótesis de comparaciones múltiples relacionadas con estos efectos, tanto si se cuenta con una simple variable dependiente asociada con cada una de las medidas repetidas como si se cuenta con más de una.

Some approximate solutions for split-plot designs with arbitrary covariance matrices.

Palabras clave: Diseños multivariados de medidas repetidas, esfericidad multimuestral, medias ponderadas y noponderadas, procedimientos robustos y poderosos.

Clasificación AMS: 62K10, 62J1

* Departamento de Psicología. Universidad de Oviedo. Plaza Feijoo, s/n. 33003 Oviedo. e-mail:gvallejo@sci.cpd.uniovi.es.

– Recibido en diciembre de 1996.
– Aceptado en mayo de 1998.

1. INTRODUCCIÓN

En las Ciencias del Comportamiento y campos metodológicos afines destaca el impulso que a lo largo de los últimos años ha recibido el empleo de los diseños de medidas repetidas y, sobremanera, los diseños split-plot o diseños que implican anidar las unidades experimentales respecto a los niveles de una o más variables y cruzarlos respecto a los niveles de otra u otras. Bajo el cumplimiento de los supuestos de normalidad conjunta multivariada, homogeneidad de las matrices de dispersión e igualdad de las varianzas correspondientes a las diferencias entre las diferentes medidas repetidas, dichos diseños han sido analizados tradicionalmente por medio del modelo mixto univariado de Scheffé (1956). A su vez, cuando se incumple alguno de los dos últimos supuestos, e inclusive ambos a la vez, lo usual es hacer uso, tanto del enfoque univariado con los grados de libertad corregidos como del enfoque multivariado; sobre todo, teniendo en cuenta que si el diseño está equilibrado, ambos enfoques son relativamente robustos a la violación de los supuestos referidos (Huynh, 1978; Keselman y Keselman, 1990; Keselman, Lix y Keselman, 1993; Rogan, Keselman y Mendoza, 1978). En estos casos, la elección entre el enfoque univariado con los grados de libertad ajustados o el enfoque multivariado descansa, amén de otras cuestiones, en consideraciones de potencia (Davidson, 1972). Por ejemplo, asumiendo que disponemos de moderados tamaños de muestra, ligeras desviaciones en el patrón de esfericidad conllevan una mayor potencia del enfoque univariado que del multivariado correspondiente; por el contrario, la situación se va invirtiendo paulatinamente a medida que nos desviamos del patrón de esfericidad requerido.

Sin embargo, como ha sido puesto de relieve en diversos trabajos de simulación, ambos enfoques analíticos dejan de ser robustos cuando el diseño carece del equilibrio adecuado y las matrices de varianzas y covarianzas poblacionales no son combinables por ser heterogéneas. En concreto, Keselman (1993) encuentra que el procedimiento de ajustar los grados de libertad por el valor de la desviación del patrón de esfericidad produce un sesgo considerable a la hora de verificar las hipótesis correspondientes a los efectos principales de la parte intra del diseño. Por su parte, Timm (1975) nos dice que otro tanto ocurre con la significación de dichos efectos cuando se contrastan mediante el enfoque multivariado. Finalmente, tampoco conviene pasar por alto que diversos autores (Belli, 1988; Huynh y Feldt, 1976; Keselman *et al.*, 1993) han descubierto que las tasas de error de Tipo I concernientes a la interacción también se encuentran seriamente distorsionadas bajo ambos enfoques. Así pues, cuando el diseño de medidas parcialmente repetidas no esté adecuadamente balanceado y el supuesto de homogeneidad de las matrices de dispersión se incumpla, conviene ser muy cuidadosos tanto con las inferencias realizadas en relación con la igualdad de las respuestas dadas a lo largo del tiempo, como con las inferencias referidas a la hipótesis que especifica que las diferencias entre los grupos no dependen de las condiciones de observación consideradas; ya que como se ha resaltado, éstas descansan, en buena medida, en el cumplimiento del susodicho supuesto. Como destaca Wilcox (1987), las pruebas paramétricas convencionales basadas en las distribuciones t o F pueden resultar seriamente afectadas cuando la homogeneidad de las varianzas no se satisface, sobre todo, cuando el número de unidades experimentales difiere de un grupo a otro. En estos casos, las pruebas estadísticas clásicas se pueden convertir, bien en excesivamente conservadoras, o bien en excesivamente liberales, con tasas de error de Tipo I por debajo del 0.01 ó por encima del 0.70 para un nivel de significación nominal del 0.05 (Keselman y Keselman, 1988; Milligan, Wong y Thompson, 1987; Wilcox, Charlin y Thompson, 1986). Para abordar el problema especificado se han intentado varios remedios, no obstante, la solución más habitual consiste en hacer uso del arreglo Behrens-Fisher y en aproximar los grados de libertad mediante el procedimiento de Welch (1938, 1947, 1951). Dicha solución fue aplicada inicialmente por Welch a los diseños de dos o más grupos al azar, y muchos años después, Algina y Olejnik (1994) la han extendido a los diseños factoriales, mientras que Keselman, Carriere y Lix (1993, 1995) lo han implementado en el ámbito de los diseños de medidas repetidas y en el ámbito de los diseños factoriales no ortogonales; todo ello al amparo de la versión generalizada que del procedimiento de Welch y de la posterior mejora de James (1951,1954) realizó Johansen en 1980. En el contexto de los diseños de medidas repetidas otro de los remedios disponibles, y a juicio de Algina (1994) y Algina y Oshima (1994, 1995) muy recomendable, es el enfoque de la aproximación general de Huynh (1978), así como subsiguientes modificaciones efectuadas en el mismo.

2. EL ENFOQUE DE WELCH-JAMES

Por lo que se refiere a la primera solución o enfoque de la aproximación multivariada de los grados de libertad desarrollado por Johansen (1980), en adelante enfoque de Welch-James, resaltar que los descubrimientos empíricos encontrados a lo largo de los últimos años en el contexto de los diseños univariados ponen de relieve la robustez de este enfoque a la hora de probar si existen diferencias entre las medias de dos o más poblaciones cuando las varianzas son heterogéneas y el número de unidades experimentales difiere de un grupo a otro. Más aún, la potencia de este enfoque resulta comparable a la potencia encontrada tras aplicar las pruebas clásicas de t o F cuando el supuesto de homogeneidad de las varianzas es satisfecho (Roth, 1983; Brown y Forsythe, 1974; Hsiung y Olejnik, 1994b, Wilcox *et al.*, 1986) y, por supuesto, mucho más poderoso que éstas cuando, tanto el tamaño de los grupos como el tamaño de las varianzas difiere sustancialmente.

No obstante, conviene matizar que el enfoque Welch-James no mantiene inalterable su robustez a lo largo de todas las situaciones. En concreto, cuando existe una correlación negativa entre el tamaño de los grupos y el tamaño de las varianzas (la varianza más grande acontece en el grupo de menor tamaño, o viceversa) y los datos presentan un sesgo apreciable, el enfoque bajo consideración no controla la tasa de error de Tipo I al nivel nominal a. Cuando se registra más de una variable dependiente o a las unidades experimentales se las mide en más de una ocasión, la tasa de error de Tipo I también puede llegar a ser considerable si la razón entre el tamaño de los grupos y el número de variables dependientes o bien el número de medidas repetidas es pequeña, las matrices de varianzas y covarianzas son heterogéneas y los datos no se acomodan a una distribución normal multivariada. En las demás situaciones el enfoque de Welch-James proporciona un adecuado grado de control de la tasa de error de Tipo I; posteriormente efectuaremos algún comentario adicional sobre esta cuestión. Con todo, no conviene dejar sin aclarar que el resto de las técnicas, tanto si son de naturaleza paramétrica como de naturaleza no paramétrica, participan de las mismas debilidades y ofrecen un grado de generalización mucho más restringido que el enfoque bajo consideración.

Seguidamente, se expone la modelización del enfoque Welch-James en el ámbito de los diseños split-plot. Para ello, inicialmente, nos ceñiremos al caso en que las $q(k, \ldots, q)$ respuestas recogidas a partir de las $n(i, \ldots, n)$ unidades muestrales independientes estén agrupadas de acuerdo con los $p(j, \ldots, p)$ niveles de una variable de clasificación. Para una situación como la descrita, el modelo lineal general con N unidades experimentales puede escribirse como sigue:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$$

donde $\boldsymbol{Y}$ es una matriz de respuestas de orden $N \times q$, $\boldsymbol{X}$ es la matriz de diseño de rango pleno de orden $N \times p$, $\boldsymbol{B}$ es una matriz de parámetros no aleatorios de orden $p \times q$ y $\boldsymbol{E}$ es una matriz de errores aleatorios de orden $N \times q$. Si denotamos por $\varepsilon_i'(\varepsilon_{ij1}, \ldots, \varepsilon_{ijk})$ el vector de errores aleatorios correspondiente a la unidad ith, se asume que cada subvector de errores es $N_q(\mathbf{0}, \boldsymbol{\Sigma}_j)$. El hecho de que la forma de $\boldsymbol{\Sigma}_j$ dependa de j indica que todos los vectores de errores aleatorios no tienen la misma matriz de varianzas y covarianzas, $\boldsymbol{\Sigma}$, lo que implica que las matrices no son combinables.

En términos sustantivos las hipótesis de interés del diseño split-plot o diseño de medidas parcialmente repetidas son las siguientes:

1. ¿Existe interacción entre las variables entre e intra del diseño?
2. ¿Difieren entre sí los diferentes grupos de tratamiento?
3. ¿Tienen todas las respuestas el mismo efecto?

Para contrastar dichas hipótesis haremos dos cosas: Por un lado, utilizaremos la prueba estadística de Welch-James, la cual, de acuerdo con Johansen (1980), puede expresarse como sigue:

$$\boldsymbol{T}_{W-J} = (\boldsymbol{R}\bar{\boldsymbol{y}})'(\boldsymbol{R}\boldsymbol{P}\boldsymbol{R}')^{-1}(\boldsymbol{R}\bar{\boldsymbol{y}})$$

donde $\bar{\boldsymbol{y}}$ es un vector de orden $pq \times 1$ obtenido tras concatenar verticalmente las medias de $\bar{\boldsymbol{y}}_j$, $\boldsymbol{R}\,(\boldsymbol{R} = \boldsymbol{C}' \otimes \boldsymbol{A}')$ es una matriz de contrastes cuyo orden depende de la hipótesis que estemos probando y $\boldsymbol{P}$ es una matriz diagonal de bloques de orden $pq \times pq$ con el jth bloque igual a $\hat{\Sigma}_j / n_j$ $(\hat{\Sigma}_1/n_1, \ldots, \hat{\Sigma}_p/n_p)$. Según Johansen (1980, p.86), el estadístico T_{W-J} dividido por una constante, c, se distribuye aproximadamente como el estadístico F con grados de libertad ν_1 (rango de la matriz $\boldsymbol{R}$) y $\nu_2 = \nu_1\,(\nu_1+2)/3A$. Las constantes c y A valen, respectivamente

$$c = \nu_1 + 2A - \frac{(6A)}{(\nu_1+2)}$$

y

$$A = \frac{1}{2}\sum_{j=1}^{p} \left\{ tr\left[\boldsymbol{PR}'\,(\boldsymbol{RPR}')^{-1}\boldsymbol{RQ}_j\right]^2 + \left[tr\,(\boldsymbol{PR}'\,(\boldsymbol{RPR}')^{-1}\boldsymbol{RQ}_j\right]^2 \right\} \Big/ (n_j - 1)$$

donde tr denota el operador traza y $\boldsymbol{Q}_j$ es una matriz diagonal de bloques de orden $pq \times pq$ con el jth bloque igual a una matriz de identidad de orden $q \times q$ y el resto ceros.

Por otro lado, expresaremos todas las hipótesis expuestas anteriormente mediante una adecuada elección de la matriz de contrastes $\boldsymbol{R}$ y también en términos de los parámetros de la matriz $\boldsymbol{B}$ que sigue:

$$\boldsymbol{B} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1q} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2q} \\ \mu_{p1} & \mu_{p2} & \cdots & \mu_{pq} \end{pmatrix}$$

Concretando aún más, la hipótesis nula que afirma que las diferencias entre los niveles de la variable de tratamiento no dependen de los niveles de la variable intra considerados, viene dada por

$$H_{0_1}: \quad \begin{pmatrix} \mu_{11} - \mu_{12} \\ \cdots \\ \mu_{1,q-1} - \mu_{1q} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{p1} - \mu_{p2} \\ \cdots \\ \mu_{p,q1} - \mu_{pq} \end{pmatrix}$$

o, bien

$$H_{0_1}: \quad \boldsymbol{R\mu} = \boldsymbol{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes \boldsymbol{A}'$ es una matriz de orden $(p-1)(q-1) \times pq$, $\boldsymbol{C}'$ es una matriz de coeficientes de orden $(p-1) \times p$ que determina los elementos de μ a incluir en la hipótesis nula, $\boldsymbol{A}$ es una matriz de orden $q \times (q-1)$ propia de las situaciones multivariadas que permite generar hipótesis entre los diferentes parámetros de respuesta, μ

es un vector de parámetros de orden $pq \times 1$ y $\mathbf{0}$ es un vector nulo de cuyo orden es $pq \times 1$. Las matrices $\boldsymbol{C}'$ y $\boldsymbol{A}$ adoptan la forma que sigue:

$$\boldsymbol{C}' = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}; \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{pmatrix}$$

La H_{0_1} se rechaza al nivel α si

$$T_{w-J}/c \ > \ F_{\alpha}; \nu_1, \nu_2 \qquad (\nu_1 = p-1 \times q-1)$$

donde $T_{w-J}/c \ \sim \ F(\nu_1, \nu_2)$.

De resultar la interacción significativa la hipótesis nula de ausencia de diferencias entre los grupos viene dada por

$$H_{0_2}: \quad \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1q} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2q} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{p1} \\ \mu_{p2} \\ \vdots \\ \mu_{pq} \end{pmatrix}$$

o, simplemente

$$H_{0_2}: \quad \boldsymbol{R\mu} \ = \ \mathbf{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes \boldsymbol{A}'$ es una matriz de orden $(p-1)\, q \times pq$, $\boldsymbol{C}'$ tiene la misma forma que en el caso anterior y $\boldsymbol{A}$ es una matriz de identidad de orden $q \times q$.

En ausencia de interacción la hipótesis nula de igualdad de los grupos se reduce a:

$$H^*_{0_2}: \quad \sum_{k=1}^{q} \mu_{1k}/q = \sum_{k=1}^{q} \mu_{2k}/q = \ \cdots \ = \sum_{k=1}^{q} \mu_{pq}/q$$

o, bien

$$H^*_{0_2}: \quad \boldsymbol{R\mu} \ = \ \mathbf{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes \boldsymbol{a}'$ es una matriz de orden $(p-1) \times pq$, $\boldsymbol{C}'$ mantiene la forma aludida y $\boldsymbol{a}$ es un vector de unos de orden $q \times 1$. En ambos casos la hipótesis nula referida a los niveles de la variable *entre* se rechaza al nivel α si

$$T_{w-J}/c \ > \ F_{\alpha}; \quad \nu_1, \nu_2$$

Finalmente, la hipótesis nula de igualdad de las respuestas u ocasiones de observación también puede abordarse teniendo en cuenta la significación o no de la interacción. En el primer caso, dicha hipótesis puede expresarse como:

$$H_{0_3}: \quad \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{pmatrix} = \begin{pmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{pmatrix} = \cdots \begin{pmatrix} \mu_{1q} \\ \mu_{2q} \\ \vdots \\ \mu_{pq} \end{pmatrix}$$

o, simplemente

$$H_0: \quad \boldsymbol{R\mu} = \boldsymbol{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes \boldsymbol{A}'$ es una matriz de orden $p(q-1) \times pq$, $\boldsymbol{C}'$ es una matriz de identidad de orden $p \times p$ y $\boldsymbol{A}$ es una matriz de contrastes de orden $q \times (q-1)$.

En el segundo caso, esto es, de no existir interacción la hipótesis de igualdad de las respuestas viene dada por

$$H^*_{0_3}: \quad \sum_{j=1}^{p} \mu_{j1}/p = \sum_{j=1}^{p} \mu_{j2}/p = \cdots = \sum_{j=1}^{p} \mu_{jq}/p$$

o, bien

$$H^*_{0_3}: \quad \boldsymbol{R\mu} = \boldsymbol{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes \boldsymbol{A}'$ es una matriz de orden $(q-1) \times pq$, $\boldsymbol{c}'$ es un vector de unos de orden $1 \times p$ y $\boldsymbol{A}$ es una matriz de contrastes de orden $q \times (q-1)$. Bajo cualquiera de las dos situaciones especificadas, la hipótesis nula referida a la igualdad de los niveles de la variable intra se rechaza al nivel α si

$$T_{W-J}/c > F_{\alpha}; \quad \nu_1, \nu_2$$

donde $T_{W-J}/c \sim F(\nu_1, \nu_2)$.

Antes de concluir con este apartado conviene advertir que el enfoque descrito sólo permite probar la hipótesis nula de igualdad de las respuestas concerniente a medias no ponderadas. El investigador interesado en probar la hipótesis nula referida a medias ponderadas puede utilizar la solución multivariada que del problema Behrens-Fisher nos ofrecen Nel y van der Merwe (1986).

3. EXTENSIÓN DEL ENFOQUE WELCH-JAMES A SITUACIONES MULTIVARIADAS

Con las modificaciones oportunas de las matrices $\boldsymbol{C}'$ y $\boldsymbol{A}$ puede manejarse cualquier tipo de diseño que implique medidas repetidas y que utilice una sola variable dependiente. Sin embargo, a nadie se le escapa que existen bastantes situaciones que si bien se acomodan a un diseño de medidas repetidas, éstas son esencialmente de naturaleza multivariada.

Considérese el diseño multivariado de medidas parcialmente repetidas esquematizado en la figura 1 tomada de Vallejo y Menéndez (1997), en el cual hay un factor de agrupamiento p $(j = 1, \ldots, p)$, un factor de observaciones repetidas q $(k = 1, \ldots, q)$ y las respuestas dadas por el ith sujeto a lo largo de q medidas repetidas bajo cada una de las r variables de medida. De este modo, bajo esta disposición las q primeras columnas corresponden a la primera variable dependiente, las q segundas a la segunda variable dependiente y las $q-$ésimas la rth variable dependiente.

Trat.	Suj.	VD_1	VD_2		VD_r
A_j	S_i	$\mathbf{y}'_{ij1} = \left[y^{(1)}_{ij1}\, y^{(2)}_{ij1}\, \ldots\, y^{(q)}_{ij1}\right]$	$\mathbf{y}'_{ij2} = \left[y^{(1)}_{ij2}\, y^{(2)}_{ij2}\, \ldots\, y^{(q)}_{ij2}\right]$	$\ldots$	$\mathbf{y}'_{ijr} = \left[y^{(1)}_{ijr}\, y^{(2)}_{ijr}\, \ldots\, y^{(q)}_{ijr}\right]$

Figura 1. Disposición de los datos del diseño multivariado de medidas parcialmente repetidas.

Para un diseño como el esquematizado el modelo lineal multivariado con N unidades experimentales puede ser escrito como sigue:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$$

donde $\boldsymbol{Y} = N \times qr$ es la matriz de respuestas, $\boldsymbol{B} = p \times qr$ es la matriz de parámetros, $\boldsymbol{X} = N \times p$ es la matriz de diseño de rango pleno y $\boldsymbol{E} = N \times qr$ es la matriz de errores aleatorios. Si denotamos por $\varepsilon'_i = 1 \times qr$ el vector de errores aleatorios correspondiente al sujeto ith, es asumido que:

$$\boldsymbol{\varepsilon}'_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_j)$$

donde la matriz de covarianzas $\Sigma_j = qr \times qr$ es una matriz definida positiva. El hecho de que la forma Σ dependa de la población de la que se hayan extraído las n_j unidades, supone que no todos los vectores de errores aleatorios ε tienen la misma matriz Σ y, por ende, son matrices heteroscedásticas.

Bajo esta nueva situación, la construcción de la matriz $\boldsymbol{R}$ para probar las hipótesis del modelo es algo más complicada; no obstante, tal cometido se puede manejar relativamente bien si se prosigue con una notación similar a la utilizada en el caso univariado. Para ilustrar lo dicho se comienza probando la hipótesis nula que especifica que las diferencias entre los grupos no dependen de las condiciones de observación consideradas en el conjunto de las r variables dependientes examinadas simultáneamente, esto es

$$H_{0_1}: \quad \begin{pmatrix} \mu_{11}^{(1)} - \mu_{12}^{(1)} \\ \vdots \\ \mu_{1q-1}^{(2)} - \mu_{1q}^{(2)} \\ \vdots \\ \mu_{1q-1}^{(r)} - \mu_{1q}^{(r)} \end{pmatrix} = \begin{pmatrix} \mu_{21}^{(1)} - \mu_{22}^{(1)} \\ \vdots \\ \mu_{2q-1}^{(2)} - \mu_{2q}^{(2)} \\ \vdots \\ \mu_{2q-1}^{(r)} - \mu_{2q}^{(r)} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{p1}^{(1)} - \mu_{p2}^{(1)} \\ \vdots \\ \mu_{pq-1}^{(2)} - \mu_{pq}^{(2)} \\ \vdots \\ \mu_{pq-1}^{(r)} - \mu_{pq}^{(r)} \end{pmatrix}$$

o, bien

$$H_{0_1}: \quad \boldsymbol{R\mu} = \mathbf{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes (\boldsymbol{A}_q \otimes I_r)'$ es una matriz de orden $(p-1)(q-1)r \times pqr$, $\boldsymbol{C}'$ es una matriz de coeficientes de orden $(p-1) \times p$ que determina los elementos de μ a incluir en la hipótesis nula, $\boldsymbol{A}$ es una matriz de coeficientes de orden $q \times (q-1)$ que nos permite generar hipótesis entre los diferentes niveles de la variable intra, $\boldsymbol{I}_r$ es una matriz de identidad de orden $r \times r$ y μ es un vector de parámetros de orden $pqr \times 1$. La H_{0_1} se rechaza al nivel α si

$$T_{W-J}/c \; > \; F_\alpha; \quad \nu_1, \nu_2 \qquad [\nu_1 = (p-1)(q-1)r]$$

donde $T_{W-J}/c \sim F(\nu_1, \nu_2)$. El cálculo de los grados de libertad correspondientes al error se efectúa igual que en el caso univariado descrito con anterioridad.

La ausencia de diferencias entre los grupos puede ser abordada desde un doble punto de vista en función de que la interacción sea o no sea significación. En el primer caso,

$$H_{0_2}: \quad \begin{pmatrix} \mu_{11}^{(1)} \\ \mu_{12}^{(1)} \\ \vdots \\ \mu_{1q}^{(1)} \\ \vdots \\ \mu_{1q}^{(r)} \end{pmatrix} = \begin{pmatrix} \mu_{21}^{(1)} \\ \mu_{22}^{(1)} \\ \vdots \\ \mu_{2q}^{(1)} \\ \vdots \\ \mu_{2q}^{(r)} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{p1}^{(1)} \\ \mu_{p2}^{(1)} \\ \vdots \\ \mu_{pq}^{(1)} \\ \vdots \\ \mu_{pq}^{(r)} \end{pmatrix}$$

o, alternativamente

$$H_{0_2}: \quad \boldsymbol{R\mu} = \mathbf{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes (\boldsymbol{A}_q \otimes \boldsymbol{I}_r)$ es una matriz de orden $(p-1)qr \times pqr$, $\boldsymbol{C}'$ tiene la misma forma que en el caso anterior, $\boldsymbol{A}_q$ e $\boldsymbol{I}_r$ son matrices de identidad de orden $q \times q$ y $r \times r$, respectivamente.

A su vez, en el segundo caso, esto es, de no existir interacción

$$H_{0_2}^{*}: \quad \begin{pmatrix} \sum_{k=1}^{q} \mu_{1k}^{(1)}/q \\ \sum_{k=1}^{q} \mu_{1k}^{(2)}/q \\ \vdots \\ \sum_{k=1}^{q} \mu_{1k}^{(r)}/q \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{q} \mu_{2k}^{(1)}/q \\ \sum_{k=1}^{q} \mu_{2k}^{(2)}/q \\ \vdots \\ \sum_{k=1}^{q} \mu_{2k}^{(r)}/q \end{pmatrix} = \cdots = \begin{pmatrix} \sum_{k=1}^{q} \mu_{pk}^{(1)}/q \\ \sum_{k=1}^{q} \mu_{pk}^{(2)}/q \\ \vdots \\ \sum_{k=1}^{q} \mu_{pk}^{(r)}/q \end{pmatrix}$$

o, alternativamente

$$H_{0_2}^{*}: \quad \boldsymbol{R\mu} = \boldsymbol{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes (\boldsymbol{a}_q \otimes \boldsymbol{I}_r)'$ es una matriz de orden $(p-1)r \times pqr$, $\boldsymbol{C}'$ e $\boldsymbol{I}_r$ adoptan la misma forma que en el caso de arriba, pero aq es un vector de unos de orden $q \times 1$. En ambos casos la H_0 se rechaza al nivel α si

$$T_{W-J}/c \;>\; F_{\alpha}; \quad \nu_1,\, \nu_2$$

donde $T_{W-J}/c \;\sim\; F(\nu_1,\, \nu_2)$.

Por último, presentamos la prueba de la hipótesis nula multivariada de igualdad de las condiciones de observación promediadas a través de los grupos. De nuevo dicha hipótesis puede abordarse desde una doble perspectiva en función de que la interacción resulte o no significativa. En el primer caso,

$$H_{0_3}: \quad \begin{pmatrix} \mu_{11}^{(1)} \\ \mu_{21}^{(1)} \\ \vdots \\ \mu_{p1}^{(1)} \\ \vdots \\ \mu_{21}^{(r)} \\ \vdots \\ \mu_{p1}^{(r)} \end{pmatrix} = \begin{pmatrix} \mu_{12}^{(1)} \\ \mu_{22}^{(1)} \\ \vdots \\ \mu_{p2}^{(1)} \\ \vdots \\ \mu_{22}^{(r)} \\ \vdots \\ \mu_{p2}^{(r)} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{1q}^{(1)} \\ \mu_{2q}^{(1)} \\ \vdots \\ \mu_{pq}^{(1)} \\ \vdots \\ \mu_{2q}^{(r)} \\ \vdots \\ \mu_{pq}^{(r)} \end{pmatrix}$$

o, simplemente

$$H_{0_3}: \quad \boldsymbol{R\mu} = \boldsymbol{0}$$

donde $\boldsymbol{R} = \boldsymbol{C}' \otimes (\boldsymbol{A}_q' \otimes \boldsymbol{I}_r)$ es una matriz de orden $p(q-1)r \times pqr$, $\boldsymbol{C}'$ es una matriz de identidad de orden $p \times p$, $\boldsymbol{A}_q$ es una matriz de contrastes de orden $q \times (q-1)$ e $\boldsymbol{I}_r$ es una matriz de identidad de orden $r \times r$.

En el segundo caso, esto es, si no existe interacción entre los niveles de la variable *entre* y las condiciones de observación en el conjunto de las variables dependientes consideradas simultáneamente, tenemos

$$H_{0_3}^{*}: \quad \begin{pmatrix} \sum_{j=1}^{p} \mu_{j1}^{(1)}/p \\ \sum_{j=1}^{p} \mu_{j1}^{(2)}/p \\ \vdots \\ \sum_{j=1}^{p} \mu_{j1}^{(r)}/p \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{p} \mu_{j2}^{(1)}/p \\ \sum_{j=1}^{p} \mu_{j2}^{(2)}/p \\ \vdots \\ \sum_{j=1}^{p} \mu_{j2}^{(r)}/p \end{pmatrix} = \cdots = \begin{pmatrix} \sum_{j=1}^{p} \mu_{jq}^{(1)}/p \\ \sum_{j=1}^{p} \mu_{jq}^{(2)}/p \\ \vdots \\ \sum_{j=1}^{p} \mu_{jq}^{(r)}/p \end{pmatrix}$$

o, simplemente

$$H_{0_3}^{*}: \quad \boldsymbol{R\mu} = \boldsymbol{0}$$

donde $\boldsymbol{R} = \boldsymbol{c}' \otimes (\boldsymbol{A}' \otimes \boldsymbol{I}_r)$ es una matriz de orden $(q-1)r \times pqr$, $\boldsymbol{c}'$ es un vector de unos de dimensión $1 \times p$, $\boldsymbol{A}_q$ es una matriz de contrastes de orden $q \times (q-1)$ e $\boldsymbol{I}_r$ es una matriz de identidad de dimensión $r \times r$. En ambos casos la hipótesis nula se rechaza al nivel α si

$$T_{W-J}/c > F_{\alpha}; \quad \nu_1, \nu_2$$

donde $T_{W-J}/c \sim F(\nu_1, \nu_2)$.

4. EL ENFOQUE DE LA APROXIMACIÓN GENERAL

Como ha sido anticipado, el procedimiento multivariado de ajustar los grados de libertad que acabamos de describir también adolece de limitaciones, y de hecho debiera ser utilizado con ciertas precauciones cuando el tamaño de muestra es reducido, sobremanera, en lo referido a la significación de la interacción. En estas circunstancias los trabajos de simulación llevados a cabo por Algina (1994), Algima y Oshima (1994, 1995) y Coombs y Algina (1996), han puesto de relieve que tanto el procedimiento de la aproximación general mejorada (IGA), como el de la corrección de la aproximación general mejorada (CIGA), proporcionan un mejor control de la tasa de error Tipo I para los efectos principales que el procedimiento Welch-James. Los procedimientos *IGA* y *CIGA* representan tan sólo variaciones del enfoque de la aproximación general (*GA*) desarrollado por Huynh (1978) para contrastar medias ponderadas en los diseños de medidas parcialmente repetidas. Por lo que se refiere a la interacción, los resultados empíricos de Algina y Oshima (1994) y de Keselman, Carriere y Lix (1993) conducen a conclusiones similares para ambos enfoques. Keselman *et al.* (1993) encuentran que el procedimiento Welch-James tiende a ser liberal cuando el tamaño de muestra es reducido. En concreto, y a tenor de los hallazgos de estos autores, para que la tasa de error Tipo I de los efectos principales no exceda de 0.075 cuando el valor nominal vale 0.05, se requiere que la razón entre n_j y $q-1$ sea de 2 cuando $q = 4$ y de 1.7 cuando $q = 8$; para la interacción la razón requerida es de 3 cuando $q = 4$ y de 3.9 cuando $q = 8$. Dichas razones están referidas a datos normales, para datos muestreados desde distribuciones sesgadas las razones deben ser ligeramente superiores. A su vez, Algina y Oshima (1994) encuentran que cuando la razón entre n_j y $q-1$ es pequeña y la matriz de dispersión se aproxima al patrón de esfericidad, el enfoque *GA* tiende a ser conservador; por el contrario, dicho enfoque se vuelve ligeramente liberal a medida que la matriz de dispersión se aleja del patrón de esfericidad. Los efectos referidos se acentúan cuando se trata comprobar la significación de la interacción y conforme los datos se apartan de la normalidad.

El enfoque de la aproximación general desarrollado por Huynh (1978), permite extender el trabajo iniciado por Box en 1954 a todas aquellas situaciones en las cuales, además de incumplirse el supuesto de esfericidad, también se incumple el supuesto de homogeneidad de las matrices de varianzas y covarianzas. De acuerdo con el enfoque *AG* propuesto por Huynh, la hipótesis nula referida a la ausencia de diferencias entre las medias ponderadas de la variable intra se rechaza si $F > \hat{b}F_{\alpha}; \hat{h}', \hat{h}$. Donde $\hat{b}$, $\hat{h}'$ y $\hat{h}$ son constantes generalmente desconocidas, pero que pueden estimarse a partir de la matriz de varianzas-covarianzas muestral, $\hat{\boldsymbol{\Sigma}}$, como sigue:

$$\hat{b} = \frac{N - p\,\mathrm{tr}\left(\boldsymbol{C}^{*'}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^{*}\right)}{\sum_{j=1}^{p}(n_j-1)\,\mathrm{tr}\left(\boldsymbol{C}^{*'}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^{*}\right)}$$

$$\hat{h}' = \frac{(\mathrm{tr}\left(\boldsymbol{C}^{*'}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^{*}\right)^2}{\mathrm{tr}\left(\boldsymbol{C}^{*'}\hat{\boldsymbol{\Sigma}}\boldsymbol{C}^{*}\right)^2}$$

y

$$\hat{h} = \frac{\left[\sum_{j=1}^{p}(n_j-1)\,\mathrm{tr}\left(\boldsymbol{C}^{*'}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)\right]^2}{\sum_{j=1}^{p}(n_j-1)\,\mathrm{tr}\left(\boldsymbol{C}^{*'}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)^2}$$

donde $\boldsymbol{C}^{*'}$ es una matriz de contrastes ortonormalizados de orden $(q-1)\times q$ y $\hat{\boldsymbol{\Sigma}} = N^{-1}\sum_{j=1}^{p} n_j\hat{\boldsymbol{\Sigma}}_j$.

A su vez, la nulidad de la otra hipótesis afectada por la ausencia de esfericidad multimuestral, esto es, la referida a la ausencia de interacción entre las variables del diseño de medidas parcialmente repetidas, se rechaza de acuerdo con el enfoque GA si $F > \hat{c}F_{\alpha};\ \hat{h}'',\ \hat{h}$. En este caso, la obtención de los estadísticos $\hat{c}$ y $\hat{h}''$ se complica ligeramente; no obstante, también pueden obtenerse a partir de las matrices de varianzas-covarianzas muestrales como sigue:

$$\hat{c} = \frac{N-p}{p-1}\frac{\operatorname{tr}\boldsymbol{G}\hat{\boldsymbol{S}}}{\sum_{j=1}^{p}(n_j-1)\operatorname{tr}\left(\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)}$$

y

$$\hat{h}'' = \frac{(\operatorname{tr}\boldsymbol{G}\hat{\boldsymbol{S}})^2}{\operatorname{tr}(\boldsymbol{G}\hat{\boldsymbol{S}})^2}$$

donde es una matriz diagonal de bloques de orden $pq \times pq$, estando los jth's bloques de la diagonal principal conformada por $\hat{\boldsymbol{\Sigma}}_1/n_1, \hat{\boldsymbol{\Sigma}}_2/n_2, \ldots, \hat{\boldsymbol{\Sigma}}_p/n_p$ y los de las diagonales secundarias por ceros, y $\boldsymbol{G}$ una matriz de bloques de orden $pq \times pq$ con los bloques de la diagonal principales tomando el valor de $n_j(1-n_j/N)(\boldsymbol{I}-\mathbf{11}'/q)$ y los de las diagonales secundarias $-N^{-1}(\boldsymbol{I}-\mathbf{11}' \div q)\,n_j n_j'$.

Para aquellas situaciones en las que el supuesto de esfericidad multimuestral esté próximo a ser satisfecho, Huynh (1978) ha señalado que los estadísticos $\hat{h}$ y $\hat{h}'$ tienden a subestimar el valor de sus parámetros correspondientes. En los casos en los que esto suceda, Huynh recomienda efectuar una mejora en el enfoque GA y utilizar el enfoque IGA mediante la sustitución de $\hat{h}$ por $\tilde{h} = \tilde{\eta}/\tilde{\delta}$, donde

$$\begin{aligned}\tilde{\eta} = {} & \sum_{j=1}^{p}\frac{(n_j-1)^3}{(n_j+1)(n_j-2)}\left\{n_j\left(\operatorname{tr}\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)^2 - 2\operatorname{tr}\left(\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)^2 + \right.\\ & \left. + \sum_{j}\sum_{j'}\left[(n_j-1)(n_j'-1)\left(\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)\left(\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)\right]\right\}\end{aligned}$$

y

$$\tilde{\delta} = \sum_{j=1}^{h}\frac{(n_j-1)^2}{(n_j+1)(n_j-2)}\left[(n_j-1)\operatorname{tr}\left(\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)^2 - \left(\operatorname{tr}\boldsymbol{C}^{*\prime}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{C}^{*}\right)^2\right]$$

A su vez, en base a la enmienda que Huynh y Feldt (1976) realizaron del estimador $\hat{\varepsilon}$ de Greenhouse y Geisser (1959), Huynh (1978), sugiere reemplazar $\hat{h}'$ por

$$\tilde{h}' = \frac{N\hat{h}'-2}{N-p-\hat{h}'}$$

y $\hat{h}''$ por

$$\tilde{h}'' = \frac{(p-1)\left[N\hat{h}''-2(p-1)\right]}{(N-p)(p-1)-\hat{h}''}$$

En conexión con el procedimiento *IGA*, Algina y Oshima (1994) sugieren efectuar una mejora del mismo en base a la corrección que Lecoutre (1991) realizó del $\tilde{\varepsilon}$ de Huynh y Feldt (1976). De acuerdo con dicha modificación, Algina y Oshima aconsejan reemplazar $\tilde{h}'$ por

$$\tilde{h}^{*\prime} = \frac{(N-p+1)\hat{h}'-2}{N-p-\hat{h}'}$$

y $\tilde{h}''$ por

$$\tilde{h}^{*\prime\prime} = \frac{(p-1)\left[(N-p+1)\hat{h}''-2(p-1)\right]}{(N-p)(p-1)-\hat{h}''}$$

Resaltar, por último, que el enfoque de la aproximación general de Huynh (1978), así como las subsiguientes modificaciones que del mismo hemos expuesto más arriba, sirve para contrastar hipótesis nulas referidas a medias ponderadas; sin embargo, al contar con distintas unidades experimentales dentro de cada grupo parece razonable que el investigador contemple este hecho y, por ende, le interese probar hipótesis referidas a medias no ponderadas. De resultar cierto lo dicho, en esta nueva situación, además de tener que computarse los estadísticos F correspondientes a la igualdad de las respuestas y a la ausencia de la interacción entre las variables efectuando un *AVAR* no ponderado, los argumentos de los valores críticos $bF_\alpha;\ \hat{h}',\ \hat{h}$ y $cF_\alpha;\ \hat{h}'',\ \hat{h}$ deben definirse en términos de la matriz de dispersión $\mathbf{\Sigma}^*$ $\left(\mathbf{\Sigma}^* = \sum_{j=1}^{p} n_j^{-1}\mathbf{\Sigma}_j \Big/ \sum_{j=1}^{p} n_j^{-1}\right)$, en vez de la matriz $\mathbf{\Sigma}$ definida anteriormente.

5. DETERMINACIÓN DE LAS DIFERENCIAS ENTRE LOS EFECTOS DEL DISEÑO EN AUSENCIA DE ESFERICIDAD MULTIMUESTRAL

Una vez que se han detectado diferencias en los efectos del diseño es necesario determinar qué comparaciones o contrastes son los responsables de las susodichas diferencias. Para alcanzar la meta reseñada, en lo que resta de este trabajo vamos a hacer dos cosas: por un lado, ofrecer alguna prueba estadística que utilice un término de error específico para cada uno de los contrastes y, por otro, describir algún procedimiento de los muchos que existen actualmente para efectuar comparaciones, que al tiempo que mantiene controlada la probabilidad de cometer un error de Tipo I al nivel de significación elegido para la familia de contrastes, no esté exento de potencia de prueba.

Como ha sido reiterado, en aquellas investigaciones que resulte significativa la interacción lo sensato sería que el investigador concentrase todos sus esfuerzos en el análisis de ésta. Para tal cometido, los investigadores más propensos a las cuestiones de carácter aplicado recomiendan centrarse en el análisis de las diferencias entre los niveles de una variable en función de cada uno de los niveles de la otra; es decir, en el análisis de los efectos principales simples; por el contrario, los investigadores más atraídos por las cuestiones de carácter teórico recomiendan analizar exclusivamente los contrastes de la interacción. A nuestro modo de ver, si obviamos lo fácil que suele resultar la interpretación de los efectos principales simples, desde un punto de vista estrictamente teórico, el segundo enfoque resulta más apropiado. La razón es sencilla, ya que para un contraste particular la hipótesis que está siendo considerada es consistente con la prueba global del efecto de la interacción.

Actualmente, existen diversos tipos de contrastes para analizar la interacción (Boik, 1993; Timm, 1994; Gabriel, Putter y Wax, 1973), no obstante, los contrastes producto o interacción contraste-contraste resultan los más fáciles de construir y también de interpretar. Por ejemplo, en un modelo con dos factores, la interacción contraste-contraste se obtiene llevando a cabo el producto Kronecker entre dos vectores, conformando cada uno de éstos un contraste entre los niveles de ese factor principal. De este modo, en un diseño de medidas parcialmente repetidas 2×3 los contrastes producto serían tres; mientras que si el diseño hubiese sido un 3×3 los contrastes producto serían nueve. En ausencia de esfericidad multimuestral, comprobar la significación de cada uno de los contrastes producto requiere hacer uso de un estadístico que nos permita obtener el error estándar de los datos implicados en el contraste de interés. Una forma de alcanzar dicho objetivo es mediante el estadístico t' que sigue:

$$t'_{\Psi} = \frac{\left(\overline{Y}_{.jk} - \overline{Y}_{.jk'}\right) - \left(\overline{Y}_{.j'k} - \overline{Y}_{.j'k'}\right)}{\sqrt{\dfrac{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_j\,\boldsymbol{c}}{n_j} + \dfrac{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}'_j\,\boldsymbol{c}}{n_{j'}}}} \sim t\,\frac{\alpha}{2},\, \nu'$$

donde $\hat{\boldsymbol{\Sigma}}_j$ es la matriz de varianzas-covarianzas correspondiente al jth nivel de la variable entre y $\boldsymbol{c}$ es un vector de coeficientes de orden $q \times 1$ que recoge los niveles kth y k'th de la variable intra sujetos. Si bien el estadístico anterior no sigue exactamente una distribución t de Student, puede ser aproximado a la misma calculando los grados de libertad mediante el procedimiento de Satterthwaite (1946) que sigue:

$$\nu' = \frac{\left[\dfrac{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_j\,\boldsymbol{c}}{n_j} + \dfrac{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_{j'}\,\boldsymbol{c}}{n_{j'}}\right]^2}{\dfrac{\left[\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_j\,\boldsymbol{c}\right]^2}{n_j^2\,(n_j - 1)} + \dfrac{\left[\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_{j'}\,\boldsymbol{c}\right]^2}{n_{j'}^2\,(n_j - 1)}}$$

Otra vía más sencilla de alcanzar los mismos resultados, es mediante el procedimiento Welch-James desarrollado a lo largo del presente trabajo. En este caso concreto, el enfoque Welch-James no es ni más ni menos que una versión generalizada de la prueba t' expuesta más arriba. Esta nueva situación requiere usar el vector de contrastes $\boldsymbol{r}$ que sigue:

$$\boldsymbol{r} = \boldsymbol{c}'_{jj'} \otimes \boldsymbol{a}'_{kk'}$$

donde $\boldsymbol{c}'_{jj'}$ y $\boldsymbol{a}'_{kk'}$ son vectores de coeficientes que contrastan los niveles de las variables entre e intra, respectivamente.

A partir de aquí, al investigador sólo le resta utilizar alguno de los numerosos procedimientos que existen para efectuar comparaciones de medias, y que le permita mantener la tasa de error de Tipo I controlada al nivel nominal para todas las comparaciones que componen la familia de contrastes de la interacción. Pues bien, sendos trabajos de simulación llevados a cabo por Keselman y Lix (1996) y Lix y Keselman (1996), ponen de relieve que tanto el procedimiento secuencial de arriba-abajo de Holm (1979) y modificado por Shaffer (1986), como el procedimiento de abajo-arriba de Hochberg (1988) controlan adecuadamente la tasa de error de Tipo I para los contrastes de la interacción en diseños con características similares a las aquí aludidas; esto es, diseños de medidas parcialmente repetidas con matrices de varianzas y covarianzas heterogéneas y desigual número de unidades experimentales dentro de cada grupo.

A su vez, para contrastar las hipótesis referidas a las comparaciones múltiples efectuadas entre los niveles de la variable intra sujeto cuando el diseño, además de tener un número arbitrario de sujetos dentro de cada grupo, no satisface el supuesto de esfericidad multimuestral, también conviene utilizar algún estadístico que permita calcular el error estándar a partir de los datos implicados en el contraste de interés. Una forma de alcanzar dicho objetivo consiste en utilizar el estadístico t' convenientemente adaptado a esta situación como sigue:

$$t'_{\Psi} = \frac{\left(\overline{Y}_{..k} - \overline{Y}_{..k'}\right)}{\sqrt{\dfrac{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_j\boldsymbol{c}}{n_j} + \dfrac{\boldsymbol{c}'\hat{\boldsymbol{\Sigma}}_{j'}\boldsymbol{c}}{n_{j'}}}} \sim t\,\frac{\alpha}{2},\,\nu'$$

donde $\hat{\boldsymbol{\Sigma}}_j$ denota la matriz de varianzas-covarianzas para el jth nivel de la variable entre y $\boldsymbol{c}'$ es un vector de coeficientes de orden $1 \times q$ referido a los niveles kth y k'th de la variable intra. Como en el caso de la interacción, el estadístico t' no se distribuye exactamente como t de Student, pero puede aproximarse a dicha distribución muestral calculando los grados de libertad ν' mediante el procedimiento de Satterthwaite allí descrito.

Alternativamente, una forma más fácil de obtener pruebas robustas y también poderosas de las comparaciones múltiples para los niveles de la variable intra, consiste en utilizar el manido procedimiento de Welch-James; pues como el lector habrá reparado, la única modificación a lo ya dicho afecta tan sólo a la construcción del vector de contrastes $\boldsymbol{r}$. Para este caso concreto, $\boldsymbol{r} = \boldsymbol{c}'_j \otimes \boldsymbol{a}'_{kk'}$, donde $\boldsymbol{c}_j$ es un vector de unos de dimensión $p \times 1$ y $\boldsymbol{a}_{kk'}$ es un vector de contrastes referido a las diferentes comparaciones entre los niveles de la variable intra cuya dimensión es $q \times 1$.

Señalar, por último, que para la hipótesis de igualdad de las medidas repetidas, los resultados empíricos de Keselman y Lix (1996) ponen de relieve que el procedimiento protegido de Shaffer (1986) proporciona un adecuado control de la tasa de error de Tipo I, al tiempo que se manifiesta ligeramente más poderoso que el resto de los procedimientos testados. Seguidamente, vamos a efectuar una breve descripción de los procedimientos de Holm-Shaffer y Hochberg que acabamos de reseñar a lo largo de este apartado.

5.1. El procedimiento de Holm-Shaffer

Shaffer (1986) ha propuesto una versión más poderosa del procedimiento de rechazo secuencial desarrollado por Holm (1979). Bajo la modificación desarrollada por esta investigadora uno comienza ordenando por rangos los valores ρ asociados con el estadístico de la prueba, $\rho_{(1)} \leqslant \rho_{(2)} \leqslant \cdots \leqslant \rho_{(c)}$. A continuación, el valor ρ más pequeño es comparado con el valor crítico de Dunn, esto es, α/c, donde $c = q(q-1)/2$ y α denota la tasa de error a controlar; si $\rho_{(1)} \leqslant \alpha/c$ la hipótesis nula $H_{(1)}$ es rechazada, de lo contrario, detenemos el proceso y declaramos nulas todas las hipótesis, $H_{(1)}, H_{(2)}, \ldots, H_{(c)}$. Si $H_{(1)}$ es rechazada, procedemos a comparar el siguiente valor ρ más largo, $\rho_{(2)}$, aquí a diferencia del procedimiento de Holm, en vez de dividir α por el número de pares de comparaciones restantes, $c-1$, Shaffer propone dividirlo por el número de pares de comparaciones que podrían ser iguales, condicionados al rechazo de la hipótesis previa, c_2^*; si $\rho_{(2)} \leqslant \alpha/c_2^*$, la hipótesis $H_{(2)}$ es rechazada, en caso contrario, detenemos el proceso. El procedimiento continúa del modo expuesto rechazando H_k $(k = 1, 2, \ldots, c)$ si $\rho_{(k)} \leqslant \alpha/c_k^*$, dados $k-1$ rechazos previos. Por ejemplo, cuando $k = 4$ la diferencia entre pares de medias más larga es probada frente a un determinado valor crítico $(t, F, q, \ldots)$ basada en $\alpha/6$, pues $c = 4(4-1)/2 = 6$. En esta primera fase, tanto el procedimiento de Holm como el de Shaffer son plenamente coincidentes; no obstante, en la segunda etapa el procedimiento de Holm probaría el siguiente par de medias más largo basado en $\alpha/5$, mientras que el procedimiento de Shaffer verificaría dicha diferencia usando $\alpha/3$, pues el rechazo de la primera hipótesis (por ejemplo, $\mu_1 \neq \mu_4$) implica que al menos tres de los restantes pares de medias podrían aún seguir siendo iguales (por ejemplo, $\mu_1 = \mu_2$, $\mu_1 = \mu_3$ y $\mu_2 = \mu_3$ ó $\mu_2 = \mu_4$, $\mu_3 = \mu_4$ y $\mu_2 = \mu_3$). Para la tercera y cuarta comparaciones más largas, el procedimiento de Holm utilizaría $\alpha/4$ y $\alpha/3$, respectivamente; sin embargo, el procedimiento de Shaffer emplearía $\alpha/3$ en ambos casos, ya que el rechazo de la segunda hipótesis (por ejemplo, $\mu_1 \neq \mu_3$) implica que un conjunto de tres hipótesis puede seguir siendo verdadero (por ejemplo, $\mu_2 = \mu_4$, $\mu_3 = \mu_4$ y $\mu_2 = \mu_3$), a su vez, el rechazo de la tercera hipótesis (por ejemplo, $\mu_2 \neq \mu_4$) aún permitiría que un conjunto de tres hipótesis podría seguir siendo verdadero (por ejemplo, $\mu_1 = \mu_2$, $\mu_3 = \mu_3$ y $\mu_3 = \mu_4$). Finalmente, para las dos últimas comparaciones, ambos procedimientos usarían de nuevo el mismo α'; esto es, $\alpha/2$ y $\alpha/1$. En la tabla

1 tomada de Shaffer (1986), se muestra la secuencia de divisores (c_k^*) para k = 3, 4, 5, 6 y 7.

Tabla 1. *Número de posibles pares de medias que pueden permanecer iguales entre k medias con r rechazos.*

(r) número de hipótesis rechazadas

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	3	1	1																		
4	6	3	3	3	2	1															
5	10	6	6	6	6	4	4	3	2	1											
6	15	10	10	10	10	10	7	7	7	6	4	4	3	2	1						
7	21	15	15	15	15	15	15	11	11	11	11	10	9	7	7	6	5	4	3	2	1

Shaffer ha tabulado el valor de c_k^* para $k = 3, \ldots, 10$.

En el mismo trabajo de 1986, Shaffer propuso otra modificación del procedimiento de Holm. La única diferencia entre el procedimiento que acabamos de describir y la segunda modificación propuesta por Shaffer, radica tan sólo en el empleo de una prueba de igualdad de medias global. Es decir, que con esta versión protegida de Shaffer se procede a efectuar el contraste de medias una vez verificado que la hipótesis nula global es falsa. Tras el rechazo de la hipótesis nula global, lo cual implica que al menos una comparación es distinta de cero, sin que necesariamente tenga que ocurrir la diferencia entre un simple par de medias, se procede como en el caso descrito pero asumiendo el rechazo de la hipótesis más larga; de este modo, para el ejemplo especificado, si la hipótesis nula global es rechazada, la diferencia más larga sería verificada utilizando $\alpha/3$, en vez de $\alpha/6$, para las hipótesis restantes procederíamos de un modo similar a lo dicho para la versión no protegida. Por consiguiente, si la hipótesis nula global es rechazada, entonces la versión protegida de Shaffer ofrece una mayor potencia del test estadístico asociado con el par de medias que difieren más entre sí.

Si bien la autora a la que nos venimos refiriendo, ha tabulado el valor de c_k^* para distintos niveles de la variable ($k = 3, \ldots, 10$), para efectuar el análisis de los contrastes de la interacción recomienda utilizar la rutina que sigue: Si $H_{(1)}$ es rechazada utilizando el criterio de Dunn, α/c, probamos $H_{(2)}$ empleando como divisor c_2^*, donde $c_2^* = c - (p-1)(q-1)$; esto es, α/c_2^*. A su vez, si $H_{(2)}$ es rechazada, se fija $c_k^* = c_2^*$ para todo $2 \leqslant k \leqslant c - c_2^* - 1$ y asignamos un valor de $c - k + 1$ para todo $k > c - c_2^* + 1$. Por ejemplo, si tenemos un diseño factorial $A \times B$ con $p = 2$ y $q = 4$, el número total de contrastes para la interacción, c, es igual a $p^* \times q^* (p^* = p(p-1)/2$ y $q^* = q(q-1)/2)$. Así pues, de acuerdo con la primera versión del procedimiento de Shaffer, para $c = 1, \ldots, 6$, tenemos los siguientes valores de α':

$$
\begin{aligned}
\alpha'_1 &= \frac{\alpha}{c^*_1} = \frac{\alpha}{6} \\
\alpha'_2 &= \frac{\alpha}{c^*_2} = \frac{\alpha}{3} \quad \text{donde} \quad c^*_2 = 6-(2-1)\times(4-1) \\
\alpha'_3 &= \frac{\alpha}{c^*_3} = \frac{\alpha}{3} \quad \text{pues} \quad 2 \leqslant 3 \leqslant 6-3+1 \\
\alpha'_4 &= \frac{\alpha}{c^*_4} = \frac{\alpha}{3} \quad \text{pues} \quad 2 \leqslant 4 \leqslant 6-3+1 \\
\alpha'_5 &= \frac{\alpha}{c^*_5} = \frac{\alpha}{2} \quad \text{ya que} \quad 2 \leqslant 5 \leqslant 6-3+1 \quad \text{es falso. Por tanto} \quad c^*_5 = 6-5+1 \\
\alpha'_6 &= \frac{\alpha}{c^*_6} = \frac{\alpha}{1} \quad \text{ya que} \quad 2 \leqslant 6 \leqslant 6-3+1 \quad \text{es falso. Por tanto} \quad c^*_6 = 6-6+1
\end{aligned}
$$

Para la versión protegida procederíamos del mismo modo, excepto para la primera diferencia, donde utilizaríamos $\alpha/3$, en vez de $\alpha/6$.

5.2. El procedimiento de Hochberg (1988)

El procedimiento paso a paso de Hochberg (1988) es una de las técnicas secuenciales más simples de cuantas existen. Para implementar este método lo primero que hay que hacer es ordenar por rangos los diferentes valores ρ asociados con el estadístico utilizado para probar las hipótesis correspondientes a los diferentes contrastes; esto es, $\rho_{(1)} \leqslant \rho_{2)} \leqslant \cdots \leqslant \rho_{(c)}$. Bajo este enfoque se parte de la hipótesis que tiene la mayor probabilidad de ser retenida a la que tiene la menor, para ello se comienza testando la hipótesis asociada con el valor ρ más largo. Si $\rho_{(c)} \leqslant \alpha$, rechazamos todas las hipótesis y detenemos el proceso; por el contrario, si $\rho_{(c)} > \alpha$, retenemos $H_{(c)}$ y procedemos a probar $H_{(c-1)}$. Si $\rho_{(c-1)} \leqslant \alpha/2$, además de la hipótesis implicada rechazamos todas las demás; en caso contrario, $H_{(c-1)}$ es retenida. A continuación, $\rho_{(c-2)}$ es comparado con $\alpha/3$, y así se continúa el proceso hasta que alguna hipótesis resulte rechazada. De ser retenidas todas las hipótesis el proceso se finaliza comparando $\rho_{(1)}$ con α/c.

6. A MODO DE CONCLUSIÓN

En el presente trabajo se ponen de relieve algunos de los problemas más importantes a los que tienen que enfrentarse los investigadores que usan los diseños de medidas repetidas, en especial, cuando no se satisface el supuesto de esfericidad multimuestral y el número de unidades experimentales varía a lo largo de los diferentes grupos de los que consta la variable de tratamiento. A su vez, a lo largo del mismo se describen con cierto grado de detalle varias pruebas que se hallan disponibles actualmente, y cuya utilización permitiría a los investigadores solventar muchos de los problemas a los que irremediablemente les conducen los estadísticos utilizados tradicionalmente en estas circunstancias. Específicamente, mostramos cómo aplicar el enfoque Welch-James para probar hipótesis relacionadas con los efectos de los diseños univariados y multivariados de medidas repetidas. Para lograr este cometido se hace un énfasis especial en la posibilidad de que la interacción resulte significativa, aspecto éste que a pesar de resultar de capital importancia, pasa desapercibido para la mayoría de los investigadores; así como, en la forma de llevar a cabo inferencias para determinar los contrastes responsables de las diferencias obtenidas en los efectos del diseño.

Con todo, y a pesar de ofrecer un amplio abanico de posibilidades de aplicación, el enfoque Welch-James no es la panacea. En concreto, para datos distribuidos normalmente su robustez se va deteriorando a medida que la razón entre el tamaño del grupo más pequeño y el número de medidas repetidas menos uno es inferior a 1.7 ó 2, cuando se prueban hipótesis referidas a los efectos principales y a 3 ó 4, para el caso referido a la interacción. Obviamente, cuando los datos se apartan de la normalidad el deterioro de la robustez de este enfoque se deja sentir para razones inferiores a 3 ó 4, en el caso de los efectos principales y para razones inferiores a 5 ó 6, para el caso de la interacción.

Finalmente, concluimos reseñando que cuando un investigador se encuentre inmerso en alguna situación en las que se incumpla el supuesto de esfericidad multimuestral, los datos estén muestreados desde distribuciones probablemente sesgadas y el tamaño de los grupos sea reducido, lo mejor que puede hacer es reducir su nivel de significación, o bien utilizar alguno de los otros enfoques descrito en este artículo. Sobre manera, si se tiene en cuenta que en estas circustancias los estudios de simulación llevados a cabo por Algina y Oshima (1995) ponen de relieve que tanto el procedimiento *IGA*, como *CIGA*, proporcionan un mejor control de la probabilidad de cometer un error de Tipo I para los efectos principales, que el enfoque Welch-James. En lo que a la interacción se refiere, los trabajos de Algina y Oshima (1994) y de Keselman *et al.* (1993), sugieren la necesidad de seguir desarrollando pruebas alternativas. Por ejemplo, una candidata potencial podría ser la versión generalizada de la prueba de Brown y Forsythe (1974).

7. AGRADECIMIENTOS

El autor desea dejar constancia pública de su agradecimiento al evaluador anónimo, pues la nueva versión del trabajo se beneficia, tanto de sus enjundiosos comentarios como de sus acertadas sugerencias.

REFERENCIAS

[1] **Algina, J.** (1994). «Some alternative approximate tests for a split-plot design». *Multivariate Behavioral Research*, **29**, 315-384.

[2] **Algina, J.** y **Olejnik, S.F.** (1984). «Implementing the Welch-James procedure with factorial designs». *Educational and Psychological Measurement*, **44**, 39-46.

[3] **Algina, J.** y **Oshima, T.C.** (1994). «Type I error rates for Huynh's general approximation and improved general approximation tests». *British Journal of Mathematical and Statistical Psychology*, **47**, 151-165.

[4] **Algina, J.** y **Oshima, T.** (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.

[5] **Belli, G.M.** (1988). *Type I error rates of MANOVA of repeated measures under group heterogenity in unbalanced designs*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

[6] **Boik, A.J.** (1993). «The analysis of two-factor interactions in fixed effects linear models». *Journal of Educational Statistics*, **18**, 1-40.

[7] **Box, G.E.P.** (1954). «Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification». *Annals of Mathematical Statistics*, **25**, 290-302.

[8] **Brown, M.B.** y **Forsythe, A.B.** (1974). «The small sample behavior of some statistics which test the equality of several means». *Technometrics*, **16**, 129-132.

[9] **Coombs, W.T.** y **Algina, J.** (1996). «On sample size requeriments for Johansen's test». *Journal of Educational and Behavioral Statistics*, **21**, 169-179.

[10] **Davidson, M.L.** (1972). «Univariate versus multivariate test in repeated measures experiments». *Psychological Bulletin*, **77**, 446-452.

[11] **Gabriel, K.R., Putter, J.** y **Wat, Y.** (1973). «Simultaneous confidence intervals for product-type interaction contrasts». *Journal of the Royal Statistical Society*, **Serie B, 35**, 234-244.

[12] **Greenhouse, S.W.** y **Geisser, S.** (1959). «On methods in the analysis of profile data». *Psychometrika*, **43**, 161-175.

[13] **Hochberg, Y.** (1988). «A sharper Bonferroni procedure for multiple tests of significance». *Biometrika*, **75**, 800-802.

[14] **Holms, S.** (1979). «A simple sequentially rejective multiple test procedure». *Scandinavian Journal of Statistics*, **6**, 65-70.

[15] **Hsiung, T.** y **Olejnik, S.** (1994). *Type I error rates and statistical power for the James second order test and univariate F test in two-way fixed effects ANOVA models under heterocedasticity and/or normality*. Paper presented at the annual meeeting of the American Educational Research Association, New Orleans, LA.

[16] **Huynh, H.** (1978). «Some approximate tests for repeated measurement designs». *Psychometrika*, **43**, 161-175.

[17] **Huynh, H.** y **Feldt, L.S.** (1976). «Conditions under which mean square ratios in repeated measurements design have exact $F-$distributions». *Journal of the American Statistical Association*, **65**, 1582-1585.

[18] **James, G.S.** (1951). «The comparison of several groups of observations when the ratios of the population variances are unknown». *Biometrika*, **38**, 324-329.

[19] **James, G.S.** (1954). «Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown». *Biometrika*, **41**, 19-43.

[20] **Johansen, S.** (1980). «The Welch-James approximation of the distribution of the residual sum of squares in weighted linear regression». *Biometrika*, **67**, 85-92.

[21] **Keselman, H.J.** (1993). «Stepwise multiple comparisons of repeated measures means under violations of multiple sphericity». En F.M. Hope (Ed.), *Multiple Comparisons, Selection and Applications in Biometry*, 167-186. New York: Marcel Dekker.

[22] **Keselman, H.J.** y **Keselman, J.C.** (1988). «Repeated measures multiple comparison procedures: Effects of violating multisample sphericity in unbalanced designs». *Journal of Educational Statistics*, **13**, 215-236.

[23] **Keselman, H.J.** y **Keselman, J.C.** (1990). «Analysing unbalanced repeated measures designs». *British Journal of Mathematical and Statistical Psychology*, **43**, 265-282.

[24] **Keselman, H.J.** y **Lix, L.M.** (1995). «Improved repeated measures: Stepwise multiple comparison procedures». *Journal of Educational and Behavioral Statistics*, **20**, 83-99.

[25] **Keselman, H.J., Carriere, M.C.** y **Lix, L.M.** (1993). «Testing repeated measures hypotheses when covariance matrices are heterogeneous». *Journal of Educational Statistics*, **18**, 305-319.

[26] **Keselman, H.J., Carriere, M.C.** y **Lix, L.M.** (1995). «Robust and powerful nonorthogonal analyses». *Psychometrika*, **60**, 395-418.

[27] **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, GA.

[28] **Lecoutre, B.** (1991). «A correction for the approximate test in repeated measures designs with two or more independent groups». *Journal of Educational Statistics*, **16**, 371-372.

[29] **Lix, L.M.** y **Keselman, H.J.** (1995). «Approximate degrees of freedom tests: A unified perspective on testing for mean equality». *Psychological Bulletin*, **117**, 547-560.

[30] **Lix, L.M.** y **Keselman, H.J.** (1996). «Interaction contrasts in repeated measures designs». *British Journal of Mathematical and Statistical Psychology*, **49**, 147-162.

[31] **Milligan, G.W., Wong, D.W.** y **Thompson, P.A.** (1987). «Robustness properties of nonorthogonal analysis of variance». *Psychological Bulletin*, **101**, 464-470.

[32] **Nel, D.G.** y **van der Merwe, C.A.** (1986). «A solution to the multivariate Behrens-Fisher problem». *Communications in Statistics-Theory and Methodology*, **15**, 3719-3735.

[33] **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.

[34] **Roth, A.J.** (1983). «Robust trend tests derived and simulated: Analogs of the Welch and Brown-Forsythe tests». *Journal of the American Statistical Association*, **78**, 972-980.

[35] **Satterthwaite, F.E.** (1946). «An approximate distribution of estimates of variance components». *Biometrics*, **2**, 110-114

[36] **Shaffer, J.P.** (1986). «Modified sequentially rejective multiple test procedures». *Journal of the American Statistical Association*, **81**, 826-831.

[37] **Timm, N.H.** (1975). *Multivariate Analysis with applications in Education and Psychology*. Monterey, CA: Brooks/Cole.

[38] **Timm, N.H.** (1994). *Analysis of interactions: Another look*. Paper presented at the annual meeting of the American Educational Association, New Orleans, LA.

[39] **Vallejo, G.** y **Menéndez, I.A.** (1997). «Una comparación de enfoques alternativos para el análisis de diseños multivariados de medidas repetidas». *Psicothema*, **9(3)**, 657-666.

[40] **Welch, B.L.** (1938). «The significance of the difference between two-mean when the population variances are unequal». *Biometrika*, **29**, 350-362.

[41] **Welch, B.L.** (1947). «The generalization of "Student's" problem when several different population variances are involved». *Biometrika*, **34**, 28-35.

[42] **Welch, B.L.** (1951). «On the comparison of several mean values: An alternative approach». *Biometrika*, **38**, 330-336.

[43] **Wilcox, R.R.** (1987). «New designs in analysis of variance». *Annual Reviev of Psychology*, **38**, 29-60.

[44] **Wilcox, R.R., Charlin, V.L.** y **Thompson, K.L.** (1986). «New Monte Carlo results on the robutness of the ANOVA F, W and F^* Statistics». *Communication in Statistics-Simulation and Computation*, **15**, 933-943.

ENGLISH SUMMARY

SOME APPROXIMATE SOLUTIONS FOR SPLIT-PLOT DESIGNS WITH ARBITRARY COVARIANCE MATRICES

G. VALLEJO SECO
J.R. ESCUDERO GARCÍA
Universidad de Oviedo*

This paper describes in some detail alternative types of analyses for unbalanced split-plot designs when the assumption of multisample sphericity is violated. Specifically, by adapting a multivariate approximate degrees of freedom approach developed by Johansen (1980) and a revised improved general approximation procedure based on Huynh (1978) it is shown how to obtain robust and powerful analysis to test within-subjects main effect and the between x within interaction, as well as multiple comparison hypotheses related to these effects, so much if is counted with a single dependent variable associated with each repeated measurement as if is counted with multiple dependent variables.

Keywords: Multivariate repeated measures designs, multisample sphericity, weighted and unweighted means, robust and powerful procedures.

AMS Classification: 62K10, 62J1

* Departamento de Psicología. Universidad de Oviedo. Plaza Feijoo, s/n. 33003 Oviedo. e-mail:gvallejo@sci.cpd.uniovi.es.

– Received December 1996.
– Accepted May 1998.

The design of repeated measures is frequently used in the investigations carried out in social, behavioral and health sciences. If the following assumptions are met: multivariate normality, homogeneity of dispersion matrices, equality of the variances corresponding to the differences between repeated measures (circularity or sphericity assumption) and independence among the observations, these designs have been traditionally analyzed by means of Scheffé's (1956) univariate mixed model. In this model, subjects (blocks) are random and treatments are fixed.

If the sphericity assumption is not met, which happens in most psychological data, the investigator has two options to analyze. On one hand, you can use an univariate mixed model with the degrees of freedom adjusted by means of some of the numerous existent Box type correctors (Lecoutre, 1991) and, on the other hand, a multivariate model. This last approach has the advantage of allowing that the variance-covariance matrix adopt any structure., Although it requires, in addition the accomplishment of the assumptions of multivariate normality and homogeneity of covariance, that the number of experimental units be bigger, or at least the same as the one of repeated measures. If the exposed conditions are satisfied, potency is used for the election of the univariate approach with the corrected degrees of freedom or the multivariate approach. For example, Davidson (1972) demonstrates that with moderate sample sizes and the covariance matrix lightly deviated of the required sphericity pattern, the univariate mixed model is always more powerful than the corresponding multivariate approach. The situation is gradually inverted as the dispersion matrix deviates of the required sphericity pattern.

When the multisample sphericity assumption is violated and group sizes is unequal, several investigators (Belli, 1988; Huynh and Feldt, 1976, Keselman, Carriere and Lix, 1993) wonder if both procedures are robust, mainly, if the number of experimental units differs from a group to another of the design. To deal with the specified problem, along the present work different procedures are described. Specifically, we show how to apply the Welch-James approach to prove hypothesis related with within-subjects main effect and the between *x* within interaction, so much for split-plot designs containing a single dependent variable as for split-plot designs containing multiples dependent variables. To achieve this, a special emphasis is made in the two following aspects:

a) In the hypothesis testing depending on the model (additive or interactive). Although in presence of interaction analytic strategy should be different from that when it is not present, most of investigators do not realize this question and in fact almost all of them testing their hypotheses according to the postulates of an additive model.

b) In the form to accomplish inferences to determine the contrasts responsible for the differences obtained between the levels and combinations of levels of the unbalanced partially repeated measures designs involving heterogeneous covariance matrices. In order to reach this last goal several multiple comparison procedures are presented in this paper.

However, in spite of offering many application possibilities, the approach Welch-James is enough sensitive to the assumption violations when the group sizes is not sufficiently large. In short, Keselman *et al.* (1993) have shown that the critical factor in determining the robustness of procedure for univariate repeated measures designs in which covariance matrices are heterogeneous and group sizes are unequal is the ratio of the smallest group size to the number of repeated measurements minus one. Particularly, for data with normal distribution their robustness deteriorates as the reason between the size of the smallest group and the number of repeated measures minus one is smaller than 1.7 or 2, when you try to prove hypothesis referred to the main effects. In the case referred to interaction, the deterioration occurs for reasons smaller than 3 or 4. Obviously, when the data are obtained from skewed distributions, the deterioration of the robustness of this approach appears for reasons smaller than 3 or 4, in the case of the main effects and it for reasons smaller than 5 or 6, in the case of interaction. Finally, we conclude pointing out that when an investigator is in some situation in which the assumption of multisample sphericity is not met, the data are to sample from non-normal distributions and the size of the groups is reduced, the best thing to do is to reduce its significance level. It can also be used some of the other approaches described in this article. Especially, if one bears in mind that in these circumstances, the simulation studies carried out by Algina and Oshima (1995) show that the Improved General Approximation and Corrected Improved General Approximation procedures provide a better control of the probability of committing Type I errors for the main effects than the Welch-James approach. With respect to interaction, the works of Algina and Oshima (1994) and of Keselman *et al.* (1993) suggest the necessity to continue developing alternative tests: for example, one possibility is to replace Welch-James procedure at the multivariate version of Brown and Forsythe (1974) procedure. Anyhow, further studies are needed.