ARTÍCULO TÉCNICO

Análisis Estadístico de Medidas Repetidas* I. Se cumple el supuesto de un patrón de varianza y covarianza homogéneas

Astrid Gutiérrez de G.[1] y
Carlos Manrique P.[2]

ABSTRACT

Title: **Statistical analysis of repeated measurements I. The assumption of homogeneous variance-covariance matrix is met**

This paper describes the statistical methodology for the analysis of information from experiments with repeated measurements on the same experimental unit under the assumption of homogeneous variance – covariance matrix.
The statistical model, with a structure of analysis of variance as split plot, is illustrated with data from an experiment of animal immunity. A «conservative» F-test is proposed; it allows to decide, without a previous examination of data, the adequacy of the univariate analysis of variance or the need of more rigurous procedures.
In addition, the sphericity test under the nule hypothesis (Ho) of no correlation and equal variance of orthogonal components, indicates the type of statistical tests required for testing treatments and period effects, which are discussed in part II of this topic.

Key Words: statistical methods, variance –covariance matrix

RESUMEN

En este trabajo se describe la metodología estadística para el análisis de la información proveniente de experimentos con medidas repetidas sobre la misma unidad experimental cuando se cumplen los supuestos de homogeneidad de la matriz de varianza - covarianza.
El modelo estadístico, con una estructura de análisis de varianza de parcelas divididas, es ilustrado con datos de un experimento de inmunidad animal. Se presenta la prueba F «conservadora», la cual permite decidir sin un examen previo de los datos, si el análisis univariado es el adecuado, o se requiere de un procedimiento más riguroso.
De otro lado, la prueba de esfericidad bajo la hipótesis nula (Ho) de no correlación e igual varianza de los componentes ortogonales, indica el tipo de pruebas estadísticas requeridas para probar el efecto de tratamientos y períodos los cuales serán discutidos en la parte II de este tópico.

Palabras Claves: métodos estadísticos, matriz varianza- covarianza.

INTRODUCCIÓN

Los investigadores del área agropecuaria frecuentemente conducen experimentos que involucran datos en cada una de varias unidades experimentales (i.e. plantas, animales, tubos de ensayo), bajo diferentes estímulos o tratamientos (i.e. fertilizantes, dietas, drogas) y cuyo efecto se mide a través del tiempo (i.e. días, meses). En otros casos, la información se genera cuando cada uno de varios tratamientos (i.e. drogas, niveles de riego) se aplican secuencialmente a la misma unidad experimental.

En general, el arreglo de medidas repetidas difiere del de parcelas divididas en que los niveles de uno o más factores no pueden ser asignados aleatoriamente a las unidades experimentales, tal como podría ser el tiempo de lectura, o en otros casos, los niveles de riego.

Bajo estas circunstancias, los errores correspondientes a las unidades experimentales pueden tener o no una matriz de varianza-covarianza homogénea (Milliken et al, 1986; Gill et al, 1971).

El análisis estadístico para el caso más simple de medidas repetidas, puede tener entonces un enfoque univariado, con estructura de parcelas divididas dado que el modelo y el análisis tienen dos partes, una para los sujetos (unidades experimentales) y otra para los períodos de tiempo (Steel y Torrie, 1986), o un enfoque multivariado, cuando se tiene una matriz de varianza-covarianza heterogénea, condición que puede afectar la validez de las pruebas F para el factor repetitivo.

Aunque en la práctica se observa muchas veces que la correlación entre dos períodos declina con diferencias grandes en el tiempo, se propone el análisis de varianza univariado porque la estructura e interpretación es familiar y no requiere de un manejo matricial complejo. A diferencia del método multivariado, su estadística F está definida si el número de respuestas excede los grados de libertad dentro de grupos o tratamientos. Por esta razón, Geisser y Greenhouse (1958) investigaron el análisis de varianza de perfiles, bajo el supuesto de residuales, con distribución multinormal y una matriz de covarianza general y propusieron una prueba F conservadora (basada en la máxima reducción de grados de libertad), para las pruebas de hipótesis de diferencias en respuesta y paralelismo cuyo verdadero valor α no exceda un valor de significancia específico. Proponen los autores un factor de ajuste E, a los grados de libertad, cuando el estadístico F, excede el valor crítico tradicional pero no excede el valor F tabulado encontrado con los grados de libertad conservadores.

1. Coordinación Regional Pecuaria, ICA, C.I. Tibaitatá, Apartado 151123 Eldorado, Bogotá, Colombia.
2. Programa Nacional de Biometría CORPOICA. C.I. Tibaitatá, Apartado 240142 Las Palmas. Bogotá, Colombia.

En el análisis de medidas repetidas, una condición suficiente para que las pruebas F del análisis de varianza sean válidas es que las matrices de covarianza de $\underline{\varepsilon}_{ij}$ y $\underline{\delta}_i$ denotadas por

$$Cov(\underline{\varepsilon}_{ij})=\Sigma \quad \text{y} \quad Cov(\underline{\delta}_i)=\Lambda$$

tengan una forma de simetría compuesta, implicando que las variables aleatorias están igualmente correlacionadas y tienen varianzas iguales. La simetria compuesta y la independencia de los errores son casos especiales de la condición H-F (Huynh y Feldt, 1970). Sin embargo, esta condición permite varianzas y correlaciones desiguales, siempre y cuando se ajuste el numerador y el denominador de los grados de libertad que pueden ayudar a que los niveles de protección de las pruebas F univariadas se ajusten a sus valores nominales.

La condición de H-F se evalúa aplicando la prueba de esfericidad (Anderson, 1958) a un grupo de componentes ortogonales, bajo la hipótesis nula (Ho) que los componentes ortogonales no están correlacionados y tienen varianza igual. Cuando esta hipótesis se rechaza, se utilizan las pruebas multivariadas o se hacen las correcciones antes mencionadas de los grados de libertad a las pruebas F univariadas.

En este trabajo se considera el análisis de medidas repetidas con un enfoque univariado, cuando se cumple el supuesto de un patrón de varianzas y covarianzas iguales. Se ilustra el análisis estadístico con los datos de la variable «recuento leucocitario», tomados del experimento que evalúa el posible efecto inmunosupresor de la vacuna contra la peste porcina clásica, los cuales fueron suministrados por el Programa de Biotecnología Animal de CORPOICA.

Análisis Estadístico

El modelo que describe la variable respuesta y considera los dos tamaños de las unidades experimentales es el siguiente:

$$Y_{ijk} = \mu+\tau_i+\delta_{ij} + p_k+(\tau p)_{ik}+\varepsilon_{ijk}$$

donde:

$$\mu+\tau_i+\delta_{ij}$$

es la parte del modelo que considera las unidades experimentales (animales) donde μ representa el promedio general, τ_i el efecto del tratamiento, δ_{ij} es el error asociado a las unidades experimentales.

La segunda parte del modelo:

$$p_k+(\tau p)_{ik}+\varepsilon_{ijk}$$

considera los períodos de tiempo analizados donde p_k es el efecto del período,

$$(\tau p)_{ik}$$

la interacción período por tratamiento y $\underline{\varepsilon}_{ijk}$ representan los errores asociados a dichos períodos. Se asume que

$$\delta_{ij} \; N(O,\sigma_\delta^2) \; y \; \varepsilon_{ijk} \; N(O,\sigma_\varepsilon^2).$$

Para el caso de medidas repetidas, el análisis de varianza de parcelas divididas es el adecuado bajo supuestos más generales, que contemplan ciertas formas de las matrices de covarianza $\underline{\varepsilon}_{ij}$ y δ_i denotadas por

$$Cov(\underline{\varepsilon}_{ij}) = \Sigma \;\; \text{y} \;\; Cov(\delta_i) = \Lambda$$

como es la condición de simetría compuesta (Huynh y Feldt, 1970). Dicha condición se expresa como:

$$\Sigma=\sigma^2 \begin{bmatrix} 1 & \rho & . & . & . & \rho \\ \rho & 1 & . & . & . & \rho \\ . & . & . & . & . & . \\ \rho & \rho & . & . & . & 1 \end{bmatrix}$$

La cual implica que las variables aleatorias están igualmente correlacionadas y tienen varianzas iguales.

Huynh and Feldt (1970) han demostrado que su condición necesaria y suficiente dada sobre $\underline{\varepsilon}$ para la medidas repetidas se cumple si:

$$C\,\Sigma\;C' = \lambda\, I$$

donde C es cualquier submatriz semiortogonal (p-1) x p de la matriz Helmert ortogonal.

$$\begin{bmatrix} \frac{j'}{\sqrt{p}} \\ C \end{bmatrix}$$

y λ es uno de los parámetros p + 1 que definen el patrón tipo H. j' es el vector 1xp de unos.

Prueba de Huynh-Feldt para Patrón de Covarianza Tipo H.

La prueba para evaluar el patrón de covarianza tipo H es de una matriz de covarianza esférica (Mauchly, 1940) cuya estadística es:

$$W=\frac{(p-1)^{p-1}(CSC')}{(tr\;CSC')^{p-1}}$$

y

$$X^2=-\left[v-\frac{2p^2-3p+3}{6(p-1)}\right] lnW$$

tiene la distribución Ji cuadrado con f = ½p(p-1)-1 grados de libertad cuando v es grande y Σ tiene el patrón tipo H. Anderson (1984), encontró los momentos de W y demostró que para p = 3, X^2 tiene una distribución exacta Ji-cuadrada con f=2 grados de libertad.

W se puede calcular con la siguiente expresión alterna:

$$W\;\frac{(p-1)^{p-1}\;S\;(j'S^{-1}j)}{p\,(trS-j'\,Sj/p)^{p-1}}$$

la cual indica la invarianza de W bajo la escogencia de la matriz C (Morrison, 1990).

Resultados y Discusión

En el experimento de referencia, 20 cerdos de similar condición fueron aleatoriamente asignados a dos grupos (10 animales cada uno). Los animales en uno de los grupos recibieron la vacuna problema y los otros actuaron como grupo control. Las lecturas del recuento leucocitario aquí analizadas se realizaron en cada uno de los animales con intervalos de dos días durante 5 períodos post-vacunación.

En este caso de medidas repetidas es importante conocer, además del efecto de tratamiento, el efecto que el tiempo tiene sobre la variable de referencia, aunque la validez de las comparaciones dentro de animales (entre períodos), además del supuesto de varianzas iguales, depende de la uniformidad de las correlaciones entre datos para cualquier par de perìodos; lo que estadísticamente significa una matriz de varianza - covarianza homogénea. Dicha estructura se puede examinar mediante la prueba de esfericidad aplicada a componentes ortogonales, la cual tiene una distribución Ji-cuadrada. Sin embargo , cuando no se tienen las facilidades de cómputo para realizarla, se puede aplicar directamente la prueba de F «conservadora» sugerida por Greenhouse y Geisser (1959).

Para los datos analizados en este trabajo, no se observó en el análisis de varianza una interacción período x tratamiento significativa (P=0.2178), ni efecto de tratamiento (P=0.8549). El efecto de período fué significativo (P=0.0027) indicando la necesidad de observar la tendencia que tuvo el recuento leucocitario a través del tiempo (Tabla 1). Los resultados obtenidos con la prueba de F «conservadora», es

Tabla 1. Análisis de varianza para el recuento leucocitario en un experimento con medidas repetidas.

Fuente de Variación	GL	CM	F	P>F ord.	P>F cons.
Tratamiento	1	1026169.00	0.03	0.8549	
Anim (Trat)= E(1)	18	29806420.11			
Período	4	51000514.00	4.50	0.0027 **	0.048 *
Tratamiento *Período	4	16756234.00	1.48	0.2178 NS	0.489 *
Residual = E(2)	72	11335407.33			
Total	**99**				
C.V.	**17**				

Prueba de Esfericidad aplicada a componentes ortogonales: Criterio de Mauchly = 0.4197.

Aprox Ji-cuadrada = 14.25 gl=9 Pr>X^2 = 0.1136 NS

* Significativo (P<.05)
** Significativo (P<.01)
NS No significativo.

Tabla 2. Promedios ($\overline{X}$) y desviación estandard (**DE**) del recuento leucocitario en los tratamientos y períodos analizados.

TRAT	n	$\overline{X}$	DE
1	50	19599.40	3665.84
2	50	19802.00	4438.26

PERIODO	n	$\overline{X}$	DE
1	20	21903.00	4727.76
2	20	20766.00	3018.00
3	20	18913.00	4634.29
4	20	18970.50	3144.19
5	20	17951.00	3457.14

TRAT	PERIODO	n	$\overline{X}$	DE
1	1	10	20526.00	4251.34
	2	10	20462.00	2446.99
	3	10	18676.00	4875.65
	4	10	19251.00	3096.05
2	5	10	19082.00	3547.54
	1	10	23280.00	4989.94
	2	10	21070.00	3610.49
	3	10	19150.00	4630.63
	4	10	18690.00	3333.15
	5	10	16820.00	3131.13

Tabla 3. Prueba de contraste para la tendencia lineal y no lineal del recuento leucocitario en los períodos analizados (1 a 5).

CONTRASTE	GL	C.M.	F	P > F
Lineal	1	188160600.49	16.60	0.0001 **
No - Lineal	3	5280485.16	0.47	0.7644 NS

** Significativo (P < .01)
NS No significativo

decir, con la máxima reducción de los grados de libertad, encontrados después de dividir los grados de libertad para el numerador y denominador de la tasa F por el número de períodos menos uno (en el ejemplo 1, 18 gl), arrojan resultados similares aunque con P>F más altas (0.048 y 0.489) para el efecto de período y la interacción tratamiento x período respectivamente.

Se puede entonces generalizar para esta prueba de F conservadora, que si el valor calculado de F es menor al valor crítico F tabulado, la hipótesis Ho de promedios iguales no se rechaza y la conclusión no se altera al reducir los grados de libertad. Al contrario, si el valor calculado de F excede el valor crítico tradicional (hay significancia estadística), se obtiene el valor F «conservador». Si la estadística F todavía excede el nuevo valor crítico, la Ho se puede rechazar. Si no la excede, se deben utilizar otros procedimientos más críticos detallados en el segundo trabajo de este tópico de medidas repetidas.

Es importante anotar que la prueba de esfericidad fué no significativa ($X^2 = 14.25$, P= 0.1136) indicando la existencia de una matriz varianza - covarianza homogénea y asegurando la validez de las pruebas F del análisis de varianza univariado. Los promedios del recuento leucocitario por tratamiento, período e interacción aparecen en la Tabla 2. La prueba de contrastes para evaluar la tendencia del recuento leucocitario indica un efecto líneal en el tiempo (P<.01) con una disminución en el recuento a través de los períodos analizados (Tabla 3). Cabe anotar, que cuando se observa una interacción significativa, se deben comparar los promedios de los períodos con cada uno de los tratamientos y hallar las tendencias si es necesario; y/o comparar los tratamientos en cada uno de los períodos, según el caso.

En este trabajo no se discuten las implicaciones biológicas de los resultados obtenidos ya que los datos analizados son parciales y tienen un enfoque metodológico - estadístico.

Conclusiones

El análisis de datos provenientes de medidas repetidas sobre la misma unidad experimental puede tener un enfoque univariado o multivariado según sea la estructura de los datos.

La prueba F - conservadora es un indicador de la validez de los resultados obtenidos, cuando no se conoce con exactitud la estructura de la matriz de varianza - covarianza.

El análisis estadístico de medidas repetidas con un enfoque univariado es similar al de parcelas divididas bajo el supuesto que pares de observaciones sobre la misma unidad experimental están igualmente correlacionadas.

La prueba de esfericidad determina si los componentes ortogonales están igualmente correlacionados y tienen igual varianza. Si es significativo, se debe acudir al análisis multivariado o realizar ajustes a las pruebas F univariadas.

BIBLIOGRAFÍA

Anderson, T.W. 1984. An introduction to multivariate statistical analysis. 2nd ed. John Wiley and Sons, Inc., New York.

Geisser, S., and Greenhouse,S.W. 1958. An extension of Box's results on the use of the F distribution in multivariate analysis. Ann. of Math. Stat. 29:885-891.

Gill, J.L. and Hafs, H.D. 1971. Analysis of repeated measurements of animals. J. Anim. Sci. 33(2):331-336.

Gill, J.L. 1986. Repeated measurement; sensitive tests for experiments with few animals. J. Anim. Sci. 63:943-954.

Huynh, H., and Feldt, L.S. 1970. Conditions under which mean square ratios in repeated measurements designs have exact F - distributions. J. Am. Stat. Assoc. 65:1582-1589.

Maychly, J.W. 1940. Significance test for sphericity of a normal n-variate distribution. Ann. of Math. Stat. 11:204-209.

Milliken, G.A., and Johnson, D.E. 1984. Analysis of Messy Data. Vol I: Designed experiments. Van Nostrand, New York.

Morrison, D.F. 1990. Multivariate Statistical Methods. 3ra ed. McGraw Hill Int.

SAS. 1991. SAS - System for linear models. 3ra ed. SAS Institute Inc. Cary, North Caroline.

Steel, R.G.D. and Torrie, J.H. 1988. Bioestadística. Principios y procedimientos. 2nd ed. (1a español). McGraw Hill, Inc. U.S.A.