

ANALISIS DE SERIES DESBALANCEADAS DE EXPERIMENTOS DE VARIEDADES

Angel Martínez Garza¹

INTRODUCCION

Es muy frecuente en los programas de mejoramiento de variedades de muchos cultivos, como ocurre en el caso de la caña de azúcar, que se proyecten series desbalanceadas de experimentos a través del espacio y del tiempo, donde el desbalance aparece, fundamentalmente, debido a los números desiguales de repeticiones de los tratamientos, en este caso las variedades. Se tiene, sin embargo, en las situaciones usuales, una "conexión" entre variedades de tal manera que es posible la estimación de cualquier contraste entre efectos de variedades; tal conexión entre variedades la proporcionan los testigos que usualmente son los mismos para todos los experimentos. Y así, aunque se tengan casos extremos de variedades que sólo aparezcan en un experimento, habiendo conexión es susceptible compararlas con el resto de las variedades.

Es el propósito de este trabajo presentar un método de análisis aproximado, de series desbalanceadas de experimentos de variedades tan frecuentes en la práctica.

¹ Profesor Investigador del Centro de Estadística y Cálculo del C.P., Chapingo, Méx.

METODO DE ANALISIS

El análisis estadístico se basa en un modelo lineal simplificado como el siguiente:

$$\bar{y}_{ij} = \mu + \tau_i + \tau_j + e_{ij}, \quad \begin{matrix} j=1,2,\dots, t, \\ i=1,2,\dots, p, \end{matrix} \quad (1)$$

donde μ es un efecto común a todas las observaciones, τ_i es el efecto debido al experimento i , τ_j es el efecto de la variedad j , e_{ij} es un término aleatorio de error y \bar{y}_{ij} es la media correspondiente a la variedad j , en el experimento i . Los errores se consideran como variables aleatorias no correlacionadas, con media 0 y varianza desconocida σ^2 .

Puesto que se tiene una situación desbalanceada, introduciremos el operador n_{ij} que toma el valor 1 si la variedad j ocurre en el experimento i , o bien el valor 0 si la variedad j no ocurre en el experimento i . Al aplicar mínimos cuadrados se obtiene el sistema de ecuaciones normales:

$$\mu: N_{..} \hat{\mu} + \sum_{i=1}^p N_{i.} \hat{\tau}_i + \sum_{j=1}^t n_{ij} \hat{\tau}_j = G,$$

$$\tau_i: N_{i.} \hat{\mu} + N_{i.} \hat{\tau}_i + \sum_{j=1}^t n_{ij} \hat{\tau}_j = B_i, \quad i=1,2,\dots, p,$$

$$\tau_j: N_{.j} \hat{\mu} + \sum_{i=1}^p n_{ij} \hat{\tau}_i + N_{.j} \hat{\tau}_j = T_j, \quad j=1,2,\dots, t,$$

donde $G = \sum_{ij} \bar{y}_{ij}$ es el gran total, $B_i = \sum_{j=1}^t \bar{y}_{ij}$ es el total para

el experimento i , $T_j = \sum_{i=1}^p \bar{y}_{ij}$ es el total para el tratamiento

j , $N_{..} = \sum_{ij} n_{ij}$ es el número total de observaciones, $N_{i.} = \sum_{j=1}^t n_{ij}$

es el número de observaciones del experimento i , y $N_{.j} = \sum_{i=1}^p n_{ij}$ es el número de observaciones del tratamiento j .

De las ecuaciones normales para tratamientos, es posible escribir:

$$\sum_{i=1}^p n_{ij} (\hat{\mu} + \hat{\eta}_i) + N_{.j} \hat{\tau}_j = T_j, \quad j=1,2,\dots, t; \quad (2)$$

por otra parte, de las ecuaciones normales para experimentos, se obtiene:

$$\hat{\mu} + \hat{\eta}_i = \frac{1}{N_{i.}} (B_i - \sum_{j=1}^t n_{ij} \hat{\tau}_j) \quad (3)$$

de modo que al substituir (3) en (2), se deriva el sistema de ecuaciones normales reducidas para tratamientos, eliminando experimentos, véase Kempthorne (1952):

$$(N_{.j} - \sum_{i=1}^p \frac{n_{ij}^2}{N_{i.}}) \hat{\tau}_j - \sum_{j' \neq j} \left(\sum_{i=1}^p \frac{n_{ij} n_{ij'}}{N_{i.}} \right) \hat{\tau}_{j'} = Y_{.j} - \sum_{i=1}^p \frac{n_{ij} Y_{i.}}{N_{i.}} = Q_j, \quad (4)$$

$$j=1,2,\dots, t.$$

Matemáticamente se demuestra que el sistema (4) de ecuaciones normales reducidas, es consistente en las $\hat{\tau}_j$, es decir, admite muchas soluciones en las $\hat{\tau}_j$. De aquí, si $\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_t$ es una solución cualquiera del sistema de ecuaciones normales reducidas, la suma de cuadrados debida a tratamientos, eliminando experimentos, $SC(TEE)$, es dada por:

$$SC(TEE) = \sum_{j=1}^t \hat{\tau}_j Q_j,$$

derivándose el análisis de varianza del Cuadro 1, donde SCE se obtiene por diferencia, restando de la suma de cuadrados total,

Cuadro 1. Análisis de varianza de series desbalanceadas de Experimentos de variedades.

-Fuentes de Variación	Grados de Libertad	Sumas de Cuadrados	Cuadrados Medios	F calculada
Experimentos Ignorando tratamientos	$p-1$	$\sum_{i=1}^p \frac{B_i^2}{N_i} - \frac{G^2}{N_{..}}$		
Tratamientos Eliminando Experimentos	$t-1$	$\sum_{j=1}^t \hat{\tau}_j Q_j$	$CMT = \frac{\sum_{j=1}^t \hat{\tau}_j Q_j}{t-1}$	$\frac{CMT}{s^2}$
Error	$N_{..} - p - t + 1$	SCE	$CME = s^2$	
Total	$N_{..} - 1$	$\sum_{j=1}^t \bar{y}_{ij}^2 - \frac{G^2}{N_{..}}$		

las sumas de cuadrados debidas a experimentos ignorando tratamientos y a tratamientos eliminando experimentos.

De aquí, para probar la hipótesis $H_0: \tau_1 = \tau_2 = \dots = \tau_t$, contra la alternativa $H_1: \tau_1 \neq \tau_2 \neq \dots \neq \tau_t$, el valor de $\frac{CMT}{s^2} F$, la "F calculada", se compara contra el valor tabulado de F, al nivel de significancia elegido, con $(t-1)$ y $(N_{..} - p - t + 1)$ grados de libertad. Si denotamos por F_{α} a este último valor, H_0 se rechaza en favor de H_1 , si $F > F_{\alpha}$, o bien, si $F < F_{\alpha}$, H_0 no se rechaza.

Separación de medias.

Si se desea comparar por ejemplo, las variedades 1 y 2 con r_1 y r_2 observaciones, respectivamente, una prueba aproximada para hacerlo se realizaría como sigue:

i) Calcúlese la diferencia $|\hat{\tau}_1 - \hat{\tau}_2|$ y el estadístico:

$$MDS = t_{\alpha, \eta} s^2 \left(\frac{1}{r_1} + \frac{1}{r_2} \right),$$

donde $t_{\alpha, \eta}$ es la t tabulada al nivel α de significancia con η grados de libertad, los del error en el análisis de varianza ($\eta = N - b - t + 1$), y s^2 , el cuadrado medio del error.

ii) Compárese $|\hat{\tau}_1 - \hat{\tau}_2|$ con la MDS aproximada, concluyendo que $\tau_1 - \tau_2$ es diferente de 0, si $|\hat{\tau}_1 - \hat{\tau}_2|$ es igual o mayor que la MDS, o bien, que $\tau_1 - \tau_2 = 0$, si $|\hat{\tau}_1 - \hat{\tau}_2|$ es menor que la MDS.

Empleo de SAS

La construcción del análisis de varianza es aritméticamente muy complicada y de hecho imposible de realizar con calculadoras de escritorio, cuando la serie de experimentos consta de 10 o más variedades diferentes, sin embargo, su cálculo empleando el paquete SAS de computación electrónica, es elemental. Se alimentan las variables correspondientes al experimento, la variedad y la característica en estudio, digamos las variables EXP, VAR y Y, respectivamente. A continuación, se hace uso del Procedimiento GLM de SAS, seguido de un enunciado de CLASSES, para definir a las variables EXP y VAR como variables de clasificación, enseguida se emplea un enunciado MODEL para definir el modelo lineal y, finalmente, se hace uso del enunciado LSMEANS para calcular las medias ajustadas. En el Cuadro 2 presenta esquemáticamente la estructura de un programa SAS para

el propósito que se acaba de indicar.

Cuadro 2. Programa SAS para el análisis de una serie desbalanceada de experimentos de variedades.

```
//SERIEEXP JØB CPOLL,A.MARTINEZ,CLASS=E
// EXEC SAS
//SYSIN DD *
DATA UNØ;
INPUT EXP VAR Y;
CARDS;
.....
.....
TARJETAS DE DATOS
.....
.....
PRØC GLM; CLASSES EXP VAR;
MØDEL Y=EXP VAR LSMEANS VAR;
/*
//
```

BIBLIOGRAFIA

Kempthorne, O. 1952. Design and analysis of experiments. John Wiley and Sons.