ESTRUCTURA DE VARIANZA Y TAMAÑO DE MUESTRA EN CONTRASTES LINEALES DE MEDIAS

LUIS GUILLERMO DIAZ M.

Profesor Asistente
Universidad Nacional

LUIS ALBERTO LOPEZ P.

Profesor Asociado
Universidad Nacional

RESUMEN. Se sugiere la estructura de varianza apropiada en diferentes casos de asignación de número de replicaciones de un tratamiento, para modelos de una vía de clasificación. También se determina el número de observaciones, con restricción de costos, para combinaciones de tratamientos.

1. INTRODUCCION

En la estimación de contrastes lineales para la comparación de medias en diseños experimentales, se asume con frecuencia que el conjunto de tratamientos en investigación tienen igual varianza y por lo general no existe ningún criterio de desagregación del número de repeticiones asignadas a cada tratamiento. En este trabajo se presentan las estructuras de varianza apropiadas para contrastes de combinaciones lineales de medias de tratamientos en los siguientes casos:

i) Muestras de tamaño desigual sin ningún criterio de asignación.

ii) Muestras de tamaño proporcional a la varianza.

iii) Muestras de tamaño proporcional a las desviaciones estándar.

Finalmente, al asumir costos variables en el estudio de tratamientos y al imponer una restricción lagrangiana sobre repetición-costos, se obtiene una expresión para el número de observaciones a la i-ésima combinación de tratamientos.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Los resultados presentados fueron desarrollados para modelos con un criterio de clasificación y estructura de medias en celdas; teniendo como base el trabajo de Cahn (1981).

2. Modelo

Si se caracteriza el modelo lineal por

$$Y_{ij} = \mu_{ij} + \epsilon_{ij}, \qquad \text{con} \qquad i = 1, \ldots, k \quad j = 1, \ldots, n_i \tag{2.1}$$

con los siguientes supuestos:

$$E(\epsilon_{ij}) = 0, \qquad E(\epsilon_{ij}^2) = \sigma_i^2; \quad \sigma_i^2 \text{ conocidas y además } \quad \epsilon_{ij} \sim N(0, \sigma_i^2)$$

$$\mu_{ij} : \text{efectos fijos de medias de tratamientos} \tag{2.2}$$

En Searle(1987), Hocking(1985) y Goldberger(1962) entre otros, se prueba por *mínimos cuadrados generalizados* que :

$$\hat{\mu}_{ij} = (X'V^{-1}X)^{-1}X'V^{-1}Y \tag{2.3}$$

donde $V = Diag(\sigma_i^2)$, y $V^{-1} = Diag(1/\sigma_i^2)$. Teniendo en cuenta el modelo (2.1) y la estimación obtenida en (2.3), se sigue que el **mejor estimador insesgado** de $\mu_{i\cdot}$ $MELI(\mu_{i\cdot})$ es $\overline{Y}_{i\cdot}$ con :

$$\overline{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, \qquad i = 1, \ldots, k \tag{2.4}$$

Si no se impone ninguna restricción en la desagregación del número de repeticiones por tratamiento, se sigue :

$$V(MELI(\mu_{i\cdot})) = \frac{\sigma_i^2}{n_i} \qquad i = 1, \ldots, k \tag{2.5}$$

3. Estructuras De Varianza Contrastes Lineales

Sea L una combinación lineal de medias de tratamientos

$$L = \sum_i \lambda_i \mu_i \tag{3.1}$$

Donde

$$MELI(L) = \sum_i \lambda_i \overline{Y}_i \tag{3.2}$$

Cuando no se tiene en cuenta ningún criterio de asignación de las observaciones a los tratamientos, se sigue que :

$$VAR(MELI(L)) = \sum_{i=1}^{k} \lambda_i^2 \frac{\sigma_i^2}{n_i} \tag{3.3}$$

Si en la asignación de las observaciones de los tratamientos, se tiene en cuenta la afijación proporcional a las desviaciones estándar; es decir:

$$\tilde{n}_i = \frac{N\sigma_i}{\sum_{i=1}^{k} \sigma_i} \tag{3.4}$$

de acuerdo a (3.2), se obtiene :

$$\begin{aligned} VAR(MELI(L)) &= \sum_{i=1}^{k} \lambda_i^2 \frac{\sigma_i^2}{n_i} \\ &= \frac{1}{N} \sum_l \sigma_l \sum_i \lambda_i^2 \sigma_i \\ &= \frac{1}{N} \sum_i \lambda_i^2 \big(\sum_l \sigma_l\big)^2 \end{aligned} \tag{3.5}$$

Cuando se asume que la desagregación de los tratamientos es proporcional a la varianza; o sea

$$n_i^* = \frac{N\sigma_i^2}{\sum_l \sigma_l^2}, \tag{3.6}$$

de (3.2) se sigue que

$$VAR(MELI(L)) = \frac{1}{N} \sum_i \lambda_i^2 \sum_l \sigma_l^2 \tag{3.7}$$

EJEMPLO

De Campos (1984) presenta los siguientes datos referentes a la producción en toneladas por hectárea de cuatro variedades de caña de azúcar.

	V_1		V_2		V_3		V_4	
	78	82	56	60	105	126	96	89
	86	80	63	82	112	94	90	91
	68	65	58	71	108	118	92	97
	77	76	70	59	121	105	97	98
	75	80	81	74	115	87	95	93
Media:	76.7		67.4		109.1		92.8	
Desviación:	6.27		9.57		12.0		3.32	

Se observa que hay una proporcionalidad en las desviaciones estándar, entonces por (3.4) los tamaños de muestra adecuados para cada variedad serían

$$\tilde{n}_1 = 8, \qquad \tilde{n}_2 = 12, \qquad \tilde{n}_3 = 15, \quad \text{y} \quad \tilde{n}_4 = 5.$$

Si además se tuviera interés en llevar a cabo la prueba de la hipótesis

$$H_0 : L = 0 \quad \text{siendo} \quad L = \mu_4 - \frac{1}{3}(\mu_1 + \mu_2 + \mu_3)$$

de (3.2) $\hat{L} = 8.4$, y de (3.5) $\hat{V}(\hat{L}) = 32.36$, teniendo finalmente que por (3.9) $t_c = 1.47$; al confrontarlo con una t de

$$\frac{(S_1^2 + S_2^2 + S_3^2 + S_4^2)}{\frac{1}{10}[(S_1^2)^2 + (S_2^2)^2 + (S_3^2)^2 + (S_4^2)^2]} = 26 \quad \text{grados de libertad,}$$

no se rechaza la hipótesis nula.

Conclusiones

Cuando se pretende hacer una desagregación del número de repeticiones por tratamiento, debe tenerse en cuenta la estructura de varianza del diseño. Se han considerado diseños asociados a modelos de una vía de clasificación con la estructura de varianza más general (desigual). Se proponen algunas expresiones para la determinación del tamaño de muestra (replicaciones), proporcional a la varianza y a los costos por el tratamiento. Para la realización de pruebas de hipótesis sobre combinaciones lineales de medias, se encontró el estadístico (3.9); los contrastes lineales son un caso especial de la combinación lineal aquí desarrolladas.

Bibliografía

Cahn, Ester S. (1981), *Comparison of Three Sampling Designs*, Amarican Statistician 35, 156-157.

Goldberg A. S. (1962), *Best Linear Unbiased Prediction in the Generalized Linear Regresión Model*, Journal of the American Statistical Association, 364-375.

De Campos H. (1984), *Estadística Aplicada á Experimentacao con Cana de Acucar*, FEALQ-USP.

Hocking, R. R. (1985), *The Analysis of Linear Models*, Brooks/Cole Publishing Company, Monterrey, California.

Searle, S. R. (1971), *Linear Models*, John Wiley, New York.

Searle, S. R. (1987), *Linear Models for Unbalanced Data*, John Wiley, New York.