Revista Colombiana de Estadística
Nº 10 - 1984

REPLICACIONES Y SUBMUESTRAS EN UN DISEÑO DE BLOQUES AL AZAR CON EFECTOS ALEATORIOS

Bernado Chávez Córdoba
Instituto Colombiano Agropecuario

Luis A. López Pérez
Profesor Asistente
Universidad Nacional

Resumen. En el presente trabajo se encontró el número de replicaciones y de submuestras en un diseño de bloques al azar, considerando una función de costo y la varianza del efecto medio total. El problema se solucionó usando la técnica de multiplicadores de Lagrange, imponiendo restricciones, en primer lugar en los costos y en segundo lugar a la varianza del efecto medio total. En ambos casos, el número de submuestras es el mismo; mientras que el número de replicaciones depende del presupuesto en el primer caso y la varianza en el segundo.

Introducción.

En algunos ensayos experimentales, se dificulta obtener la información de una unidad experimental, completa, induciendo al investigador a seleccionar muestras aleatorias, a través de las cuales obtiene la información necesaria para evaluar los resultados de las investigaciones.

Para encontrar el número de replicaciones y submuestras se han utilizado diferentes criterios, tales como: diferencias máximas entre medias de tratamientos fijando la significancia y la potencia de prueba; optimización de la información de medias de tratamientos con restricción en los costos del experimento y maximización de cualquier contraste entre tratamiendos, fijando los costos del experimento.

Los criterios anteriores no tienen en cuenta el modelo aleatorio, el número de tratamientos a aplicar y la varianza del efecto medio total, por ésta razón este trabajo propone un método de estimación del número de replicaciones y submuestras, considerando los supuestos anteriores.

Revisión de antecedentes.

En el estudio sobre el número de repeticiones y submuestras, en los diseños completamente al azar y bloques completos al azar, se han desarrollado diferentes técnicas; las cuales dan soluciones a partir de diversos supuestos.

Harris y otros (1948), partiendo de un modelo aleatorio en un diseño completamente al azar y teniendo en cuenta la hipótesis de igualdad de efectos entre tratamientos, calcula el número de repeticiones mediante la relación:

$$r = \frac{1}{1 + \frac{r\sigma^2_\tau}{\sigma^2_e}} \qquad (1)$$

donde r es el número de repeticiones de cada tratamiento, σ^2_τ es el componente de varianza de los tratamientos y σ^2_e es el componente de varianza, debido al error. Este procedimiento es iterativo y termina cuando se satisface la siguiente relación:

$$P_{Ha}\left[\frac{\frac{CMtrat}{\sigma^2+r\sigma^2_\tau}}{\frac{CMerror}{\sigma^2_e}} > \frac{\sigma^2}{\sigma^2+r\sigma^2_\tau} F_{(1-\alpha)}\right] = P_o,$$

con Po, la potencia de la prueba F, fija, CM-error y CMtrat, son estimadores preliminares. Cuando los tratamientos son fijos, el valor de r es semejante a (1), cambiando σ^2_{τ} por $\sum(\tau_j-\bar{\tau})^2$.

Tukey (1953), usando el criterio de la diferencia máxima entre dos medias de tratamientos, calcula el número de repeticiones en forma iterativa, en diseños completamente al azar y bloques completos al azar, usando la desigualdad:

$$Po = P_{Ha}\left[\frac{S^2}{S^2\text{previa}} \leq \frac{A^2 r}{q^2_{(k,\text{gle},1-\alpha/2)}\ S^2\text{previa}}\right]\ldots \qquad (2)$$

donde Po es la potencia que se desea alcanzar con la prueba, S^2 es la varianza del error, la cual no se conoce, S^2previa es un estimador preliminar de S^2, 2A longitud del intervalo propuesto por el investigador, $q_{(k,\text{gle},1-\alpha/2)}$ es el valor de las tablas de Tukey, con k tratamientos y gle grados de libertad del error.

De la solución (2) se obtiene:

$$r = \frac{S^2\text{previa}\ \ q^2_{(k,\text{gle},1-\alpha/2)}F_{(1-\alpha,\text{gle},\text{glprevia})}}{A^2}$$

Kempthorne (1967), suponiendo un modelo de blo ques al azar, con efectos fijos de tratamientos y fijando el costo del experimento por tratamientos como $Co = rC + rsCs$, donde Co es el presupuesto por tratamiento, C es el costo antes de la cosecha de una parcela, s es el número de submuestras y Cs costo asociado por unidad de muestreo dentro de la parcela. Obtiene el número de repeticiones y submuestras, maximizando la información de cada media de tratamiento $I = \frac{rs}{\sigma^2_s + \sigma^2_e}$, así

$$r = \frac{Co}{C + \sqrt{CC_s \frac{\sigma^2_s}{\sigma^2_e}}}, \qquad S = \sqrt{\frac{C\sigma^2_s}{C_s \sigma^2_e}}$$

con σ^2_s el componente de varianza debido al error de muestreo.

Martínez (1981), tomando los resultados de Kempthorne, propone estimar r y s a partir de un contraste entre efectos de tratamientos de la forma $\sum_{j=1}^{t} \lambda_j \tau_j$. Para un costo experimental Co dado, obtuvo idénticos resultados a Kempthorne.

Pérez (1984), propone minimizar una función lagrangiana de la forma:

$$\Phi(n) = V(n) + \lambda C.$$

donde $V(\hat{\theta})$ es la varianza del estimador, C es el costo fijo de enumeración de las unidades de muestreo y λ es el multiplicador de Lagrange. Obtiene los estimadores para unidades primarias, secundarias y terciarias.

Kirk (1982), fijando el nivel de significancia α, la potencia de la prueba de diferencia máxima de medios $1-\beta$, σ^2 la varianza del error y la diferencia máxima $\mu-\mu_o$, fijada por el investigador, determinó el número de replicaciones mediante la expresión:

$$r = \frac{(Z_\alpha - Z_\beta)^2}{\dfrac{(\mu - \mu_o)^2}{\sigma^2}} ,$$

donde Z_α es el valor tabulado de la distribución normal asociada con α y Z_β es el valor tabulado de la distribución normal asociado a β la probabilidad de error tipo 2.

Resultados.

Supuestos: A partir de un diseño en bloques al azar con submuestreo y considerando los efectos de tratamientos y bloques aleatorios, se presenta el modelo estadístico:

$$y_{ijk} = \mu + \beta_i + \tau_j + \varepsilon_{ij} + \delta_{ijk}$$

$i = 1,2,\ldots,r \qquad j = 1,2,\ldots,t \qquad k = 1,2,\ldots,a$

y_{ijk} es la variable aleatoria observable en el i-ésimo bloque, correspondiente al j-ésimo tratamiento de la k-ésima muestra; μ es el efecto de la media poblacional; $\beta_i \sim NI(0,\sigma^2_\beta)$, es el efecto del i-ésimo bloque; $\tau_j \sim NI(0,\sigma^2_\tau)$, es el efecto del j-ésimo tratamiento; ε_{ij} es el error experimental, el cual se supone tiene distribución normal e independiente, con media cero y varianza σ^2_e y δ_{ijk} error de muestreo, $\delta_{ijk} \sim NI(0,\sigma^2_\delta)$.

Determinación de r y n;

Criterio. Para determinar el número de repeticiones y muestras en este diseño, se minimizó, en primer lugar, la varianza total sujeta a la restricción de un presupuesto fijo y en segundo lugar se minimizó el costo total sujeto a una varianza total fija.

Notación.

Sea C_0 el costo fijo del experimento; C_1 el costo de un bloque antes de la cosecha y C_2 el costo de tomar una muestra dentro de ca da bloque. Se propone la siguiente función de costo:

$$C(a,r) = Co + rC_1 + raC_2.$$

La varianza del estimador de la media poblacional es:

$$V(\bar{y}\ldots) = \frac{1}{(rat)^2} V\left[\sum_{i=1}^{r} \sum_{j=1}^{t} \sum_{k=1}^{a} (y_{ijk} = \mu + \beta_i + \tau_j + \varepsilon_{ij} + \delta_{ijk})\right] = \frac{\sigma^2\tau}{t} + \frac{\sigma^2\beta}{r} + \frac{\sigma^2 e}{rt} + \frac{\sigma^2\delta}{rat}$$

Solución.

1. Para un presupuesto fijo $k = Co + rC_1 + raC_2$, se minimiza la varianza de $(\bar{y}\ldots)$. Sea la función

$$Q(a,r) = \frac{\sigma^2\beta}{r} + \frac{\sigma^2\tau}{t} + \frac{\sigma^2 e}{rt} + \frac{\sigma^2\delta}{rta} + \lambda\,(Co + rC_1 + raC_2 - k),$$

una función lagrangiana a minimizar con respecto a r, a y λ.

De la solución al siguiente sistema de ecuaciones

1.1. $\dfrac{\partial Q(a,r)}{\partial r} = -\dfrac{\sigma^2\beta}{r^2} - \dfrac{\sigma^2 e}{r^2 t} - \dfrac{\sigma^2\delta}{r^2 ta} + \lambda(C_1 + aC_2) = 0$

1.2. $\dfrac{\partial Q(a,r)}{\partial a} = -\dfrac{\sigma^2\delta}{rta^2} + \lambda r C_2 = 0$

1.3. $\dfrac{\partial Q(a,r)}{\partial \lambda} = Co + rC_1 + raC_2 - k = 0$

se obtuvo:

$$a = \frac{\sqrt{C_1\,\sigma^2{}_\delta}}{\sqrt{C_2(t\sigma^2_\beta + \sigma^2_\varepsilon)}}\ ,\ \lambda = \frac{1}{(k-Co)}\left[C_1\sqrt{\frac{t\sigma^2{}_\beta + \sigma^2_\varepsilon}{C_1 t}} + C_2\sqrt{\frac{\sigma^2{}_\delta}{C_2 t}}\right]^2$$

y

$$r = \frac{k - Co}{C_1 + \sqrt{\dfrac{C_1 C_2 \sigma^2{}_\delta}{\sigma^2{}_\beta + \sigma^2{}_e}}}$$

2. Para un presupuesto ilimitado y una varianza constante ν, se minimiza la función lagrangiana.

$$Q'(r,a) = Co + rC_1 + raC_2 + \lambda\left(\frac{\sigma^2{}_\beta}{r} + \frac{\sigma^2{}_\tau}{t} + \frac{\sigma^2{}_e}{rt} + \frac{\sigma^2{}_\delta}{rta} - \nu\right)$$

al solucionar el sistema de ecuaciones:

2.1. $\dfrac{\partial Q'(r,a)}{\partial a} = rC_2 - \dfrac{\lambda\sigma^2{}_\delta}{rta^2} = 0$

2.2. $\dfrac{\partial Q'(r,a)}{\partial r} = C_1 + aC_2 - \lambda\left(\dfrac{\sigma^2{}_\beta}{r^2} + \dfrac{\sigma^2{}_e}{r^2 t} + \dfrac{\sigma^2{}_\delta}{r^2 ta}\right) = 0$

2.3. $\dfrac{\partial Q'(r,a)}{\partial \lambda} = \dfrac{\sigma^2{}_\beta}{r} + \dfrac{\sigma^2{}_\tau}{t} + \dfrac{\sigma^2{}_e}{rt} + \dfrac{\sigma^2{}_\delta}{rta} - \nu = 0.$

se obtuvo:

$$a = \frac{\sqrt{C_1\sigma^2{}_\delta}}{\sqrt{C_2(t\sigma^2{}_\beta + \sigma^2{}_e)}}\ ,\ \lambda = \frac{1}{(t\nu\sigma^2{}_\tau)^2}\left[\sqrt{tC_1(t\sigma^2{}_\beta + \sigma^2_\varepsilon)} + \sqrt{C_2 t\sigma^2{}_\delta}\right]^2$$

y

$$r = \frac{t\sigma^2_\beta + \sigma^2_e}{t\nu - \sigma^2_\tau} \left[1 + \sqrt{\frac{C_2 \sigma^2_\delta}{C_1(\sigma^2_\beta + \sigma^2_e)}} \right]$$

D i s c u s i ó n.

En los diferentes métodos citados en la literatura relacionada con la determinación del número de replicaciones en esquemas de diseños completamente aleatorizados o en bloques completos al azar, se observa que en muchos de ellos se parte de la diferencia máxima entre medias de tratamientos. Al fijar el nivel de significancia α y la potencia de la prueba $1-\beta$, se determina en forma iterativa el número de replicaciones, manteniendo un número fijo de tratamientos, sin considerar un presupuesto necesario para satisfacer las necesidades experimentales.

En la determinación del número de replicaciones con un presupuesto fijo, se supone que los efectos de los tratamientos y bloques son fijos y las replicaciones no dependen del número de tratamientos.

Frecuentemente en el diseño de bloques al azar, se consideran los tratamientos como efectos aleatorios. Esto implica considerar la componente de varianza debida a tratamientos en la determinación del número de replicaciones.

Conclusiones.

Al minimizar la varianza del efecto medio total, restringida a un presupuesto fijo o viceversa, se encontró el mismo número de submuestras, en tanto que el número de replicaciones cambia atendiendo a la restricción impuesta a la función a minimizar.

Cuando se desee estimar el número de replicaciones en diseños de bloques con submuestreo y se considere el costo del experimento, es recomendable emplear los resultados de Kempthorne, siempre que se tengan efectos fijos de tratamientos, en caso contrario se usan los resultados propuestos. En ambos casos se debe tener en cuenta la varianza a minimizar, ya que siempre el número de replicaciones depende de la magnitud de la varianza.

* *

BIBLIOGRAFIA

Harris, M. Horwitz, D.G. and Mood, A.M. "On the determination of sample sizes in design experiments", J. Amer. Statist. Ass. 43 391-402. 1948.

Kemthorne, O., *Design and analysis of experiments*. John Wiley and sons, New York. 1967.

Kirk, R.E., *Experimental design, procedures for the behavioral sciences*, Wadsworth Inc. Belmont California. 1982.

Martínez, G.A., *Diseños experimentales*, Edit. C.P. 1981.

Pérez, A., "Determinación del tamaño de muestra en investigación", Revista Colombiana de Estadística, Bogotá, Nº 9, p.60-86. 1984.

Tukey, J.W., *The problem of multiple comparasions*. Ditto Princenton Univ. USA. 1953.

* * *