

## DETERMINACION DEL TAMAÑO DE MUESTRA EN INVESTIGACION

*Augusto Pérez Ordoñez*

Profesor Asistente  
Universidad Nacional.

**Resumen.** El presente artículo discute las consideraciones a tener en cuenta por los investigadores que utilizan el método estadístico para la determinación del tamaño de muestra, y señala las etapas que preceden la selección; hace énfasis en las diferencias entre el Diseño Experimental y el Muestreo de Poblaciones finitas para muestrear las unidades a estudiar.

**Abstract.** This paper discuss some considerations for researchers using statistical methodology about the determination of sample size including some discussion on the steps to take before the selection procedure, with special emphasis on the differences between Experimental Desing and Finite Population Sampling Theory.

## Introducción.

Sin incurrir en parcialidad puede decirse que el método de muestreo, tiene relación con todas las ramas de la ciencia y la técnica, y que el presente y futuro del manejo de información en el desarrollo y uso adecuado de esta técnica.

Además el amplio espectro del muestreo, permite el concurso de técnicos, científicos, políticos y gentes de toda condición; algunos fundamentados en investigaciones minuciosas, o en el sentido común; otros, sin base alguna que lo usen y difunden causando daños incalculables en la búsqueda de la verdad.

Quienes utilizan el método estadístico en sus investigaciones acuden al consultar en la fase de interpretación de resultados o en la selección del tamaño de muestra; sin comprender que la participación del estadístico debe ser desde la fase inicial, y que el uso adecuado del método requiere integrar las dos mayores preocupaciones de los investigadores a un proceso donde no puede obviarse fase alguna, y así obtener conclusiones provechosas, eficaces y verdaderas.

En este artículo se trata de fijar las pautas necesarias para la selección del número de unidades experimentales, con el objeto de contri

buir al mejor y mayor uso del muestreo y satisfacer una de las mayores inquietudes de científicos y técnicos.

## 1. Revisión de literatura.

La determinación del tamaño de muestra, es considerada por estadísticos y científicos como un aspecto fundamental en la interrelación *Teoría-práctica*, y es tratada por los primeros desde dos ángulos en función sólo si el escrito es de muestreo o de diseño experimental.

Tang (1938) asume normalidad en los errores, considera necesario estimar  $\sigma^2$ , y requerido el efecto medio de la diferencia  $(\mu_i - \mu)$ , así dadas las probabilidades de error Tipo I ( $\alpha$ ), y de error Tipo II ( $\beta$ ) encuentra el tamaño de muestra mediante la expresión

$$\emptyset_{(\alpha, \beta)} = \sqrt{\frac{2\lambda}{v}} \quad (1)$$

donde  $\lambda = \frac{v}{2\sigma^2} \sum (\mu_i - \mu)^2$ ,

$v$ : número de tratamientos

$(\mu_i - \mu)$ : desviaciones de las medias de tratamientos..

Sugiere para cada caso calcular  $\emptyset_{(\alpha, \beta)c}$ ; compa-

rar con valores tabulados por él; y mediante un proceso iterativo decidir por el número de repeticiones cuando  $\emptyset_{(\alpha, \beta)c} \leq \emptyset$  tabulado.

Harris et al (1948) obtienen un  $n$  (número de repeticiones) suponiendo normalidad y varianza común, suponen la máxima diferencia entre promedios de tratamientos  $d$ ; un estimador de la variabilidad  $s_1^2$  con  $d\delta_1$  grados de libertad, y con  $\alpha$  (nivel de significancia) y una constante  $K'$  (tabulada) obtiene:

$$n = 2[K' s_1/d]^2 (d\delta_2+1) \quad (2)$$

donde,  $d\delta_2$  = grados de libertad asociados al experimento, escogidos de tal forma que al variar  $d\delta_2$ , se obtienen valores de  $K'$  conducentes a un valor estable de  $n$ .

Desconocidos  $s_1^2$  y  $d\delta_1$ , estima la variabilidad mediante  $1/2(s_i+s_u)$  donde  $s_i$  y  $s_u$  corresponden a un rango supuesto de variabilidad.

Calcula  $s_u/s_i = c$ ; y cuando el valor de  $C$  es similar a  $\sqrt{\chi^2_{(\alpha)} / \chi^2_{(1-\alpha)}}$  los grados de libertad asociados a  $\chi^2_{(\alpha)}$  y  $\chi^2_{(1-\alpha)}$  son  $d\delta_1$ , mediante (2) encuentra  $n$  como antes.

Turkey (1953) cita los resultados obtenidos por Harris et al (1948) y propone hallar el número de repeticiones mediante:

$$n = s_1^2 q_{(\alpha, d\delta_2)}^2 F_{(1-\alpha) d\delta_1} / d^2. \quad (3)$$

Donde:

$s_1^2$ : es un estimador del error experimental

$F$ : valor tabulado

$d$ : diferencia máxima entre medias de tratamientos.

$q_{(\alpha)}$ : valor sintetizado de la diferencia media entre las diferencias parejas de tratamientos,  $d\delta_2$  y  $q_{\alpha}$  dependen del diseño y los valores de  $n$ .

Cochran (1957) al discutir el problema del número de repeticiones asociadas a un experimento, considera la diferencia entre dos tratamientos partiendo de la varianza de la diferencia estimada.

$d$ : diferencia máxima estimada entre medias de tratamientos.

$\delta$ : diferencia máxima entre medias de tratamientos.

$$v(d) = \frac{2s^2}{n} \quad v(\delta) = \frac{2\sigma^2}{n} \quad d \sim t \text{ (G.L.Error).}$$

Encuentra  $n$  de la expresión

$$n \geq 2(\sigma/\delta)^2 (t_1 + t_2)$$

donde  $t_1$  y  $t_2$  son valores tabulados.

El autor anota que el defecto es considerar a  $\delta = \sigma$ .

Además para datos arreglados en dos clases obtiene  $n$  en base al menor porcentaje de éxito y a la diferencia  $\delta$  a ser detectada entre tratamientos medida en porcentaje;  $n$  la tabula a diferentes niveles de significancia y probabilidad; y para límites de error predefinidos  $n$  lo obtiene a partir de:

$$n = \frac{K^2 (P_1 Q_1 + P_2 Q_2)}{L^2} \quad (4)$$

En donde:

$K$ : Desviación normal correspondiente a la probabilidad escogida.

$Q = (1-P)$

$L$ : longitud del intervalo de confianza.

Kempthorne (1967) afirma: "Frecuentemente no es posible o no es deseable medir la característica en el total de las unidades experimentales; o puede ser deseable *estimar* algunas características sobre la unidad completa y otras en una muestra aleatoria de la unidad". Plantea el modelo

$$y_{ijk} = m + r_i + t_j + e_{ij} + n_{ijk}$$

donde:

$m$  : media general

$r_i$  : efecto de replicación  $i = 1, \dots, r$

$t_j$  : efecto de tratamiento  $j = 1, \dots, t$   
 $e_{ij}$  : error experimental de la parcela ( $ij$ )  
 $n_{ijk}$  : error de muestreo de la observación  $k$ -ésima sobre la ( $ij$ )-ésima parcela  $k = 1, \dots, s$

$$\begin{aligned}
 E(e_{ij}) &= 0 & v(e_{ij}) &= \sigma_e^2 & e_{ij}'_s & \text{N.I.D.} \\
 & & v(n_{ijk}) &= \sigma_s^2 & n_{ijk}'_s & \text{N.I.D.}
 \end{aligned}$$

$e_{ij}$  y  $n_{ijk}$  son independientes.

Deriva un análisis de varianza para el modelo propuesto de donde obtiene  $\hat{\sigma}_e^2$  y  $\hat{\sigma}_s^2$ ; plantea una estructura de costo:

$$C_0 = n(C + s C_s),$$

donde:

$C_0$ : costo total

$C$ : costo por parcela excluyendo cosecha

$C_s$ : costo de cosecha por muestra.

Deduces  $n$  y  $s$  dado un costo fijo  $C_0$  por tratamiento maximizando la información  $n s / \sigma_s^2 + s \sigma_e^2$  de lo cual resulta:

$$n = \frac{C_0}{C + \sqrt{C C_s \frac{\sigma_s^2}{\sigma_e^2}}} \quad (5)$$

y

$$s = \sqrt{\frac{C \sigma_s^2}{C_s \sigma_e^2}} \quad (6)$$

Propone una solución intermedia entre la enumeración completa y el tamaño óptimo, donde  $n'$  (número de réplicas); el costo de enumeración completa  $C_h$ , y la información  $n'/\sigma_e^2$  y costo total, obtiene

$$n = \frac{C_o/C+C_h}{\sigma_e^2} \quad (7)$$

y propone una ecuación para la eficiencia, suponiendo enumeración total igual a  $\lambda$  muestras y sugiere tabular la eficiencia a varios valores de  $n$  y  $\lambda$  para decidir en situaciones particulares.

Kirk (1968) considera la determinación del tamaño de muestra como "uno de los problemas más complejos en el diseño experimental"; señala que la afectan, el efecto mínimo del tratamiento que el investigador está interesado en detectar, los niveles de tratamiento, la varianza del error y la probabilidad de error tipo I y tipo II; considera una analogía entre el poder de la metodología utilizada y el poder de la prueba, cita los resultados de Tang (1938), y propone para calcular el tamaño de muestra a:

$$n = \sqrt{\frac{k \sum (\mu_j - \mu)^2}{\sigma^2/n}} \quad (8)$$

Donde:

$k$ : niveles de tratamiento

$(\mu_j - \mu)$ : efecto mínimo del tratamiento que un experimentador desea detectar.

$\sigma_e^2$ : varianza del error en la población.

Luego de fijar  $\emptyset$  (poder de la prueba) mediante un proceso de ensayo y error halla  $n$ ; y afirma que dados estimadores razonables de los parámetros "el tamaño de muestra siempre deberá calcularse antes de iniciar el experimento", y dice que si los cálculos preliminares indican que el poder del ensayo es limitante, puede incrementarse aumentando  $n$  o con otro diseño que provea un estimador más preciso.

Además sugiere un procedimiento para obtener  $n$  sin acudir a  $\hat{\sigma}_e$ , asumiendo la mayor diferencia entre medias como  $C\sigma_e$  para todo  $C > 0$ ; y donde dados dos efectos de tratamientos  $\beta_j$  y  $\beta_{j'}$  iguales a  $C\sigma_e/2$  y  $-C\sigma_e/2$ , y los  $(k-2)$  efectos restantes iguales a cero, el valor:

$$\sum_{j=1}^k \beta_j^2 = \left\{ \frac{C\sigma_e}{2} \right\}^2 + \left\{ -\frac{C\sigma_e}{2} \right\}^2 = \frac{C^2\sigma_e^2}{2}$$

es mínima de tal manera que cualquier otra combinación de efectos producirá un mayor poder de la prueba, puesto que éste está en relación directa a  $\sum_{j=1}^k \beta_j^2$  así:

$$\varnothing = \frac{\sum \beta_j^2 / k}{\sigma_e^2 / \sqrt{n}} = \frac{\sqrt{(C^2 \sigma_e^2 / 2) / k}}{\sigma / \sqrt{n}} = \sqrt{n} (\sqrt{C^2 / 2k}) \quad (9)$$

Así, para un  $\varnothing$  establecido con  $C, \alpha$ , y  $(1-\beta)$  es posible hallar  $n$  mediante tablas presentadas por el autor sin conocer estimaciones de  $\sigma_e^2$ .

Kirk (1982) encuentra  $n$  dados,  $\alpha$ ,  $\beta$  y  $\sigma^2$ , la máxima diferencia  $(\mu - \mu_0)$  donde  $\mu_0$  es un valor esperado de la distribución muestral bajo  $H_0$ ; y  $\mu$  su análogo bajo  $H_1$ ; plantea que:

$$n = \frac{(Z_\alpha - Z_\beta)^2}{(\mu - \mu_0) / \sigma^2} \quad (10)$$

$Z_\alpha$  y  $Z_\beta$  son valores tabulados de la distribución normal standar; esta solución derivada por Kirk es válida para datos arreglados en dos clases.

Los escritos de Hansen et al (1953), Cochran (1953), Azorin (1969) y Sukhatme and Sukhatme (1970) consideran la selección del tamaño de muestra como parte fundamental de un proceso donde se pretende la estimación de promedios, totales y proporciones, donde se supone:

1. La existencia de:

$$A. \{X_i, P_i\}_{i=1}^N$$

donde  $X_i$  es la  $i$ -ésima unidad de muestreo y  $P_i$  la probabilidad asociada de selección.

B.  $\hat{\theta} = f(x_1, \dots, x_n)$  con densidad  $f_{\hat{\theta}}(\hat{\theta})$ , y al cual se le puede obtener  $E(\hat{\theta})$ ,  $v(\hat{\theta})$  y  $\widehat{v}(\theta)$ .

C. Una colección de muestras con una probabilidad asociada a cada muestra, y donde siempre se selecciona una muestra aleatoria de la colección definida.

2. La sección de la muestra depende fundamentalmente del plan de muestreo y del estimador seleccionado.

3. La muestra seleccionada es tal que:

$$P[|\hat{\theta} - \theta| \geq d] = \alpha$$

$d$ : error máximo permisible

$(1-\alpha)$ : nivel de seguridad en la estimación.

4. La muestra seleccionada debe producir la mejor información en términos de la varianza de  $\hat{\theta}$ , al menor costo.

5. Las derivaciones para los tamaños de muestra incluyen factores de corrección por finitud, en virtud de que la teoría del muestreo se fundamenta en la matemática de poblaciones finitas aunque algunos sugieren despreciarlos en casos particulares.

## 2. El problema y su solución.

### 1. Formulación y planteamiento del problema:

1.1. El estudio de los antecedentes sugiere las siguientes inquietudes.

1.1.1. El análisis de la situación es unilateral por parte de quienes tratan el tema de la selección del tamaño de muestra, quizá porque suponen existe la capacidad en el investigador de usarlos según sea el caso, o porque los tratadistas se restringen a su especialidad.

1.1.2. Aun cuando el investigador conoce la existencia del Diseño Experimental y el Muestreo, y en el mejor de los casos, su uso y fundamentación teórica; en la mayoría de las veces no comprende la relación entre los dos, y tampoco cuándo y cómo usarlos.

1.2. La experiencia de quienes hacemos consultoría estadística, y especialmente en muestreo, nos señala que el investigador acude a la asesoría para la interpretación de resultados, y/o, para resolver el problema del número de unidades experimentales a seleccionar, sin realizar etapas previas para que la selección de la muestra tenga sentido, y donde el concurso del consultor es indispensable.

### 2. Solución al problema planteado:

Si se acepta la importancia del método de muestreo en el desarrollo de investigaciones, y en especial, en aquellas de carácter aplicado donde se buscan las relaciones de las cosas objetivas mediante el estudio de un subconjunto del conjunto de interés; y que la relación del número de individuos no es el resultado aislado de la aplicación de una ecuación, sino una etapa fundamental de un proceso, debemos describir estas etapas y además hacer claridad en las diferencias a tener en cuenta cuando en el proceso se busca una solución mediante el uso directo de las técnicas de muestreo o de diseños experimentales.

#### Etapas del proceso:

2.1. Convencerse de las ventajas del muestreo tales como:

2.1.1. Obtención de información en forma económica y oportuna, porque el costo y el tiempo de enumerar  $n$  unidades de una población de tamaño  $N$  siempre será menor que hacer enumeración total.

2.1.2. Obtención de información más precisa porque un menor volumen de información, permite contar con personal especializado y hacer un manejo más minucioso.

2.1.3. Cuando los errores de no muestreo son muy frecuentes es más precisa la muestra que la enumeración total.

2.1.4. Existen ensayos de tipo destructivo, de control de calidad, o donde la calidad de la información requerida (p.e. estudios genéticos) no hace posible la enumeración total.

2.1.5. Siempre se estudia un subconjunto de la población; luego la enumeración completa puede considerarse una muestra grande; ahora en el caso ideal de lograr la enumeración total, el muestreo es un gran auxiliar; ejemplos contundentes son las encuestas intercensales; o cuando se pretenda estudiar características delicadas, es más conveniente hacerlo en una parte de la población.

2.1.6. Se se considera el coeficiente de variación como medida de la incertidumbre en un ensayo, es factible demostrar que su comportamiento frente al tamaño de muestra puede representarse así:

Coeficiente  
de  
variación



Tamaño de muestra

2.2. Definir el conjunto de individuos objetivo, o población motivo de la inferencia y verificar si coincide o nó sobre la población muestreada (o sobre la que se hacen las mediciones). En la población objetivo existe un conjunto de relacio

nes intrínsecas o extrínsecas a ella, y si definimos esa población como:

$$A = \{X_i, Y_i, \dots, Z_i\}_{i=1}^N$$

la selección del número de unidades experimentales dependerá de si:

- i) lo único que se pretende es encontrar un estimador  $\hat{\theta}$ .
- ii) cuantificar el efecto de  $j$  tratamiento sobre  $X$  ó  $Y$  ó  $Z$ .

En una palabra, los objetivos de la investigación han de ser los que normen las etapas posteriores.

2.3. Cuando la población es finita, debe definirse el conjunto

$$S = \{X_i, P_i\}_{i=1}^N$$

o marco de muestreo para dar estructura probabilística al muestreo, y hacer posible la inferencia; en nuestros países las limitaciones para construir marcos de muestreo constituyen el mayor obstáculo para la planificación y ejecución de muestreo; en un diseño experimental este paso puede obviarse.

2.4. Definir el grado de precisión y exactitud del estudio; la exactitud se refiere a  $d$ , o error máximo permisible entre  $\theta$  y  $\hat{\theta}$  para el

caso de estudios donde el objetivo es  $\hat{\theta}$ , o la máxima diferencia entre tratamientos para el caso del diseño experimental, y  $\alpha$  nos da la precisión, y significa el riesgo en la estimación o en el experimento.

**2.5. Especificación de las variables y los métodos de análisis y medida.**

En esta etapa se deben tener en cuenta lo siguientes:

**2.5.1. Las variables seleccionadas han de estar en correspondencia bilateral con los objetivos del estudio.**

**2.5.2. Seleccionadas,  $X_1, X_2, \dots, X_p$ , se puede aplicar cualquier método de análisis (regresión, econométrico, paramétrico o no paramétrico), obvio sin violar los supuestos respecto de la aleatoriedad, arreglo geométrico, características de las variables o comportamiento de los errores. En esta etapa puede plantearse un diseño experimental para estudiar la(s) variable(s) de interés en las unidades seleccionadas como método de análisis.**

**2.5.3. En este paso según sean los intereses del investigador se diseña un instrumento para recabar información, que puede ser un cuestionario o libro de campo.**

**2.6. Muestra preliminar:**

La mayoría de las investigaciones requieren confrontar lo planificado con la realidad; la estructura espiralada del método científico es una perfecta analogía; la muestra preliminar tiene dos sentidos, uno hacer estimaciones de varianza, requeridas por muestristas y por los especialistas en diseño para definir el tamaño de muestra, y para probar la lógica y suficiencia del instrumento de medida.

### **2.7. Definición del plan de muestreo:**

El diseño de muestreo incluye, la selección de estimadores, la determinación del tamaño de muestra y el sistema de seleccionar las unidades de muestreo.

**2.7.1.** La selección de estimadores se reduce en muestreo cuando el objetivo es  $\hat{\theta}$ , a la estimación de medias totales y proporciones es indispensable hallar  $v(\hat{\theta})$ , y su estimador, para comparar con otros esquemas de muestreo bien en términos de eficiencia relativa o de diferencia de varianzas; siempre a un plan de muestreo corresponde un tamaño de muestra y un estimador. En diseño experimental se plantearían en este momento las funciones paramétricas estimables, siempre con el objetivo de probar efectos de tratamientos.

**2.7.2.** Respecto del tamaño de muestra se deben tener en cuenta las siguientes consideraciones:

i) Dado un plan de muestreo donde se obtiene  $\hat{\theta}$  y  $n$ , dependiendo de los objetivos, se optaría por estimar  $\theta$  ó arreglar las unidades seleccionadas y plantear modelos de la forma  $Y = XB + E$ .

ii) Existirán situaciones donde se seleccionen las unidades experimentales de acuerdo sólo a los criterios establecidos por la teoría y práctica del diseño y análisis de experimentos recurriendo a seleccionar las unidades experimentales de acuerdo a las técnicas usuales de determinación de número de repeticiones, como las reportadas por la literatura o atendiendo a circunstancias específicas del diseño.

iii) En cualquier circunstancia, el tamaño de muestra debe satisfacer las siguientes condiciones:

- $n$  tal que  $P[|\hat{\theta} - \theta| > d] = \alpha$ , donde  $d$  es el error máximo permisible, bien entre  $\hat{\theta}$  y  $\theta$ , o entre medias de tratamientos.
- Debe proporcionar la mejor información en términos de varianza a menor costo.

**2.7.3.** Los esquemas de selección, su fundamentación teórica y la forma de ejecutarlos son propios de la teoría de poblaciones finitas.

De acuerdo a las deducciones de Hansen et al (1953), Cochran (1953), Azorin (1969) y Sukhatme and Sukhatme (1970), en especial a lo que

que muestreo polietápico se refiere, y obviando las correcciones por finitud en cada etapa, se deduce la siguiente situación que ilustra perfectamente la determinación del tamaño de muestra:

### Modelo propuesto para determinar $n$

Sea  $\mathcal{S}(n) = v(\hat{\theta}) + \lambda C$  donde

$v(\hat{\theta})$ : varianza del estimador

$\lambda C$ : costo de enumeración

$n$ : tamaño de muestra tal que  $\mathcal{S}(n)$  mínima, por lo cual hay dos alternativas:

- i) Dado un costo fijo minimizar varianza.
- ii) Dada una varianza fija minimizar costo.

Existirán dos problemas dependiendo si:

- i) El objetivo de la investigación se cumple al aplicar solamente las técnicas de muestreo.
- ii) Se requiere probar los efectos de tratamiento.

En el primer caso se plantea un modelo de la forma:

$$y_{ijk} = m + a_i + b_{ij} + c_{ijk} \quad [A]$$

donde  $i = 1, \dots, n_1$      $j = 1, \dots, n_2$      $k = 1, \dots, n_3$

$m$  : es un efecto general

$a_i$ : efecto de la U.P.M. (Unidad Primaria de Muestreo)

$b_{ij}$ : efecto de la unidad secundaria  $j$  en la unidad primaria  $i$ .

$c_{ijk}$ : efecto de la observación  $k$  sobre la unidad secundaria de muestreo  $j$  en la unidad primaria  $i$ .

$a_i$ ,  $b_{ij}$  y  $c_{ijk}$ : son V.A.I. (Variables Aleatorias Independientes).

Para el caso particular:

$$\hat{\theta} = \frac{\sum_{ijk} y_{ijk}}{n_1 n_2 n_3} \quad \text{y} \quad v(\hat{\theta}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3}$$

Dada una estructura de costo:

$$C_o = n_1 C_1 + n_1 n_2 C_2 + n_1 n_2 n_3 C_3 \quad (\text{A})$$

donde  $C_o$ : costo total;  $C_1$ ,  $C_2$  y  $C_3$ : el costo de enumerar una unidad primaria, secundaria y terciaria.

Si  $\Phi(n) = v(\hat{\theta}) + \lambda(C_o - n_1 C_1 - n_1 n_2 C_2 - n_1 n_2 n_3 C_3)$

los valores de  $n_1$ ,  $n_2$  y  $n_3$  que minimizan  $\Phi(n)$  se obtienen derivando parcialmente  $\Phi(n)$ , igualando a cero, así:

$$\frac{\delta(\Phi)}{\delta(n_1)} = -\frac{\sigma_1^2}{n_1^2} - \frac{\sigma_2^2}{n_1^2 n_2} - \frac{\sigma_3^2}{n_1^2 n_2 n_3} - \lambda(C_1 + n_1 C_2 + n_1 n_2 C_3) = 0 \quad (\text{B})$$

$$\frac{\delta(\Phi)}{\delta(n_2)} = -\frac{\sigma_2^2}{n_1 n_2^2} - \frac{\sigma_3^2}{n_1 n_2^2 n_3} - \lambda(n_1 C_2 + n_1 n_2 C_3) = 0. \quad (C)$$

$$\frac{\delta(\Phi)}{\delta(n_3)} = -\frac{\sigma_3^2}{n_1 n_2 n_3} - \lambda n_1 n_2 n_3 = 0. \quad (D)$$

Resultan cuatro ecuaciones con cuatro incógnitas:  $\lambda, n_1, n_2, n_3$ .

Resolviendo (A), (B), (C) y (D) se obtiene:

$$n_1 = \frac{C_0}{C_1 + n_2 C_2 + n_2 n_3 C_3}$$

$$n_2 = \frac{\sigma_2}{\sigma_1} \sqrt{\frac{C_1}{C_2}}$$

$$n_3 = \frac{\sigma_3}{\sigma_2} \sqrt{\frac{C_2}{C_3}}$$

Ahora, si  $v_0 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3}$  (A\*)

y

$$\Phi(n) = n_1 C_1 + n_1 n_2 C_2 + n_1 n_2 n_3 C_3 + \lambda \left( v_0 - \frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_1 n_2} - \frac{\sigma_3^2}{n_1 n_2 n_3} \right)$$

Al derivar  $\Phi(n)$  respecto de  $n_1, n_2, n_3$ , e igualar a cero, con (A\*) se obtiene un sistema de cuatro ecuaciones con cuatro incógnitas donde los valores  $n_1, n_2$  y  $n_3$  que los satisfacen son:

$$n_1 = \frac{n_2 n_3 \sigma_1^2 + n_3 \sigma_2^2 + \sigma_3^2}{n_2 n_3 v_0}$$

Los resultados para  $n_2$  y  $n_3$  son idénticos a los obtenidos antes.

Es importante anotar que los valores  $\sigma_1^2$ ,  $\sigma_2^2$ ,  $\sigma_3^2$  son estimados mediante un muestreo preliminar, a partir del cual se obtiene un análisis de varianza.

<i>F de variacion</i>	<i>G. de L.</i>	<i>Suma de cuadrados</i>	<i>Cuadrados medios</i>
Entre U.P.M.	$n_1 - 1$	$S_1 = \frac{\sum_i y_{i..}^2}{n_2 n_3} - C$	$M_1 = \frac{S_1}{n_1 - 1}$
U.S.M. dentro de U.P.M.	$n_1(n_2 - 1)$	$S_2 = \frac{\sum_i \sum_j y_{iji}^2}{n_3} - \frac{\sum_i y_{i..}^2}{n_2 n_3}$	$M_2 = \frac{S_2}{(n_1(n_2 - 1))}$
U.T.M. dentro de U.S.M.	$(n_3 - 1)n_1 n_2$	$S_3 = (S_1 + S_2)$	$M_3 = \frac{S_3}{(n_3 - 1)n_2 n_1}$

$$\text{TOTAL } (n_1 n_2 n_3) - 1 \quad S_T = \sum_{ijk} y_{ijk}^2 - C$$

$$\hat{\sigma}_1^2 = \frac{M_1 - M_2}{n_2 n_3} ; \quad \hat{\sigma}_2^2 = \frac{M_2 - M_3}{n_3} \quad \text{y} \quad \hat{\sigma}_3^2 = M_3$$

$$C = \left( \sum_{ijk} y_{ijk} \right)^2 / n_1 n_2 n_3 .$$

Debemos resolver:

- A. No se hace prueba de significancia porque no interesan las propiedades distribucionales y obvio las pruebas de hipótesis.
- B. De acuerdo a que el modelo es lineal y además las operaciones en que se basa la técnica (de Lagrange) son lineales, el método puede generalizarse así:

Sea la función objetivo  $\vartheta(n) = v(\hat{\theta}) + \lambda C_0$

donde

$$v(\hat{\theta}) = \frac{\sigma_3^2}{n_3} + \frac{\sigma_2^2}{n_1 n_2} + \dots + \frac{\sigma_i^2}{\prod_{i \neq j=1}^{\gamma} n_j}$$

$$C_0 = \sum_{i=1}^{\gamma} C_i n_i \quad (A^{**})$$

Se obtiene

$$\frac{\delta(\vartheta(n))}{\delta(n_i)} = 0 ; \dots ; \frac{\delta(\vartheta(n))}{\delta(n_j)} = 0 \quad j = 1, \dots, \gamma$$

(A\*\*) y las derivadas parciales de  $\vartheta(n)$  respecto de  $n_i$ , forman un sistema de  $\gamma+1$  ecuaciones con incógnitas  $\lambda, n_1, n_2, \dots$  y  $n_\gamma$  que al resolverlo resultan los valores de  $n_i$  que minimizan  $\vartheta(n)$  a un costo dado.

En forma análoga puede plantearse el método

do: dada  $v_0$  fija, se encuentran los valores que minimizan  $\theta_{(n)}$ .

Para el segundo caso se presentaría un modelo como el sugerido por Kempthorne (1967) y que es susceptible de generalizarse, de acuerdo a las fuentes de variación consideradas.

$$y_{ijk} = m + r_i + t_j + e_{ij} + n_{ijk}. \quad [B]$$

Los resultados para  $r$  y  $t$  son respectivamente (5) y (6).

Sólo debemos notar las siguientes diferencias fundamentales con el modelo anterior:

1. En [B] se pretende probar hipótesis respecto de los tratamientos, esto obliga a suponer normalidad de  $y_{ijk}$ ,  $e_{ij}$  y  $n_{ijk}$ , mientras que en el modelo [A] sólo se busca  $\hat{\theta}$ , y se hace caso omiso de la distribución de  $V$ .
2. El número de tratamientos es fijo, por tanto no se optimiza; en muestreo se consideran etapas de la selección.
3. Cuando las unidades por etapa pueden identificarse y además los factores de corrección por finitud no se pueden ignorar, las derivaciones se alteran por la presencia de los factores anotados.

4. Kempthorne (1967) solo resuelve el sistema para  $k$  (bloques) y  $s$  (unidades muestreadas) por que el número de tratamientos es fijo.

## **Conclusiones.**

1. Cuando el investigador pretenda determinar el tamaño de muestra debe además de situar este problema como una etapa del proceso indicado, contemplar las diferencias fundamentales entre muestreo y diseño experimental para resolver mejor la situación; éstas se refieren a la ausencia de tratamientos en muestreo; en muestreo las poblaciones son finitas y no interesan las propiedades distribucionales.

2. Al muestrista corresponde crear métodos de selección de unidades experimentales de acuerdo a las necesidades del diseño experimental y la investigación en general, incursionando en el diseño y fenómeno a investigar sólo para incrementar la eficiencia del método de muestreo; la teoría del muestreo polietápico, por ejemplo, puede adaptarse a la determinación del tamaño de muestra en diseño. El investigador por su parte debe plantear las hipótesis, decidir por los estadísticos y ser el guía de la investigación.

3. El diseño experimental puede considerarse como fase del muestreo cuando se usa en la etapa de selección de variables y de métodos de análisis y medida.

4. También el especialista en diseño puede colaborar con el muestrista en aspectos de análisis de varianza como el planteado por Deming (1953), citado por Cochran (1953).

5. Por las características de las operaciones se plantea la posibilidad de deducir un método general para encontrar  $n$  adaptable a cualquier plan de muestreo o diseño experimental, lo cual sería un gran aporte a la solución de este interesante problema.

\* \*

## BIBLIOGRAFIA

- Azorin, P.F., "Curso de Muestreo y Aplicaciones". Aguilar S.A. de Ediciones, 1969.
- Cochran, W.G., "Sampling Techniques". John Wiley and Sons. New York, N.Y. 1963.
- Cochran, W.G. and Cox, G.M., "Experimental Designs". John Wiley and Sons. New York, N.Y. 1957.
- Hansen, M.H., Horwitz, D.G. and Madow, L.H. "Sam-

- ple Survey Methods and Theory". John Wiley and Sons. New York, N.Y. 1953.
- Kemphorne, O., "Design and Analysis of Experiments". John Wiley and Sons. New York, N.Y. 1967.
- Kork, R.E., "Experimental design: Procedures for the Behavioural Sciences". Wadsworth, Inc., Belmont, California, 1982.
- Kirk, R.E., "Experimental design: Procedures for the Behavioral Sciences". Wadsworth, Inc., Belmont, California, 1968.
- Harris, M., Horwitz, D.G. and Mood, A.M., "On the determination of sample sizes in designing experiments". J.A.S.A. 43:391-402, 1948.
- Sukhatme, P.V. and Sukhatme, B.V. "Sampling Theory of Surveys with Applications". Iowa State University Press. Ames, Iowa, U.S.A. 1970.
- Tang, P.C. "The Power function of the Analysis of Variance Test with Tables and illustrations of their use". Stat.Res.Mem.2: 126-157. 1938.
- Tukey, J.W., "The problem of multiple comparisons". Ditto Princeton Univ., U.S.A. 1953.