matemática: Una publicación del ICM – ESPOL
2009, Vol. 7, No. 2

DISEÑO MUESTRAL ESTRATIFICADO BASADO EN LA UTILIDAD DE LA INFORMACIÓN

Salvador-Jijón Marcelo[1]

Resumen. *Tradicionalmente, el diseño muestral estratificado se hace en base al error muestral de cada estrato y a partir de éstos se calcula el error muestral total. Si se trata de minimizar el error muestral total, el resultado es el muestreo proporcional. Es práctica de los muestristas no usar este tipo de diseño, sino calcular el tamaño de la muestra en cada estrato en base a valores mínimos de errores muestrales en relación con las necesidades de la investigación. Esta metodología es intuitiva y requiere de amplia experiencia en el tema. En contraste, se propone una metodología basada en la utilidad de la información de cada uno de los estratos. Para modelar la utilidad de la información, la cual es subjetiva y variable y está en relación al conocimiento de los estratos y los objetivos de la investigación, se ha utilizado el valor de la información basado en la reducción de la entropía de Shannon (1948). Luego, de acuerdo a la utilidad de la información se plantea un problema de programación no lineal, cuya solución maximiza la utilidad total de la información para un muestreo estratificado. Por último, mediante las condiciones de Khun-Tucker, se encuentra la solución óptima al problema planteado, la cual nos permite obtener una metodología de diseño muestral estratificado no intuitiva y basada en la utilidad de la información.*

Palabras clave: Diseño muestral estratificado, utilidad, información, entropía, programación no lineal.

1. INTRODUCCIÓN

Cuando se dispone de recursos suficientes (tiempo, personal, dinero) para realizar una investigación basada en un diseño muestral estratificado, entonces se puede tener una muestra grande con una asignación proporcional a cada uno de los estratos. Pero en muchos casos, especialmente en los países no desarrollados como Ecuador, el caso anterior es más bien la excepción por lo cual generalmente se hacen investigaciones con muestras medianas o pequeñas, que se aproximan mas bien al Diseño de Experimentos. Durante muchos años hemos estado trabajando bajo esas condiciones y la solución ha sido más bien intuitiva y basada en la experiencia, es decir calcular el tamaño de la muestra en cada estrato en base a valores mínimos de errores muestrales en relación con las necesidades de la investigación. En contraste con esta metodología, propongo otra basada en la Teoría de Información de Shannon (1948) y que se expone a continuación:

2. INFORMACIÓN

En 1928, Hartley introdujo el concepto de la medición de la cantidad de información en el contexto de la Teoría de las Comunicaciones y se relaciona con el número de bits necesarios para transmitir la información. Sea X un conjunto finito de cardinalidad $|X|$. Puede ser seleccionada una sucesión de tamaño s de elementos de X. Para ilustrar el concepto supóngase que $X=\{0,1\}$ y $s=5$, eso quiere decir que una palabra $A=01111$, representa 5 bits de información. Esto puede ser medido de acuerdo a la fórmula $I = log_2 2^5 = 5$ [bits] de información. Este enfoque fue sistematizado por Hartley como se muestra a continuación:

2.1. CANTIDAD DE INFORMACIÓN DE HARTLEY

Sea X un conjunto con cardinalidad $|X|$, donde la cardinalidad de un conjunto es el número de elementos contenidos en el conjunto. Entonces, la cantidad de información de una sucesión s de elementos de **X**, con repetición, es:

$I = \ln|X|^{/s/}$ [nats] de información (2.1)

(1 nat = 1.44 bits aproximadamente)
donde /s/ es el número de elementos seleccionados para obtener información. Por otro lado, existe la necesidad de modelar el valor de la información obtenida. Para ilustrar el tema, supóngase que transmitimos 5000 bits de 0´s y a continuación 1 bit=1. Obviamente los 5000 bits no representan mucho valor de información, sino el 1 bit que está a continuación. La cantidad de información sería de 5001 bits, pero el más relevante es el bit 5001. Para modelar el valor de la información, se puede utilizar la incertidumbre, es decir mediante la probabilidad como lo hizo Shannon (1948). Por tanto, modelar el valor de la información, depende de la probabilidad asignada al conjunto X, es decir del conocimiento previo de X.

2.2. FUENTE DE INFORMACIÓN.

DEFINICIÓN. Es un par ordenado *(X,P)* donde $X = \{x_i\}$ es un conjunto finito, conocido como el

[1] *Salvador-Jijón Marcelo, Departamento de Ciencias Administrativas y Económicas, Escuela Politécnica Nacional. Quito – Ecuador. (e_mail: xsalmar@hotmail.com)*

alfabeto fuente y P la distribución probabilística de X. Denotamos la probabilidad de x_i por p_i.

2.4. VALOR DE LA INFORMACIÓN DE SHANNON (REDUCCIÓN DE LA ENTROPÍA).

DEFINICIONES. Dada una fuente de información (X,P), el valor de la información obtenida al muestrear un elemento cualesquiera x_i seleccionado de X con probabilidad p_i es:

$H(x_i) = -\ln(p_i)$ [nats] de información (2.2)

El valor de la información de un conjunto $A \subseteq X$ se define por:

$H(A) = -\ln P(A)$ [nats] de información (2.3)

2.5. TEOREMA

Dada una población finita de tamaño N, dividida en k estratos de tamaños $N_1, N_2,\ldots, N_k$, la cual se considera una fuente de información, donde la función P es la probabilidad de selección de cada uno de los elementos. Entonces, H_i, el valor de la información de Shannon del estrato i, relacionado a la probabilidad de seleccionar todos los elementos del estrato i es igual a:

$$H_i = ln\Gamma(N+1) - ln\Gamma(N_i+1) - ln\Gamma(N-N_i+1)$$

[nats] (2.4)

Prueba:

Sea p_i = probabilidad de seleccionar todos los elementos del estrato i. La probabilidad de seleccionar el primer elemento del estrato i, sin reposición es:

$\frac{N_i}{N}$; el segundo elemento: $\frac{N_i-1}{N-1}$; el tercero: $\frac{N_i-2}{N-2}$, y así sucesivamente.

Por tanto, la probabilidad de seleccionar los N_i elementos es:

$$\frac{N_i(N_i-1)(N_i-2)\ldots 1}{N(N-1)(N-2)\ldots(N-N_i+1)} =$$

$$\frac{N_i(N_i-1)(N_i-2)\ldots 1(N-N_i)(N-N_i-1)\ldots 1}{N(N-1)(N-2)\ldots(N-N_i+1)(N-N_i)(N-N_i-1)\ldots 1} =$$

$$\frac{N_i!(N-N_i)!}{N!}$$

Luego:

$$H_i = -\ln(p_i) = \ln(N!) - \ln(N_i!) - \ln((N-N_i)!) =$$

$$\ln\Gamma(N+1) - \ln\Gamma(N_i+1) - \ln\Gamma(N-N_i+1)$$

(2.4)

3. MUESTREO ESTRATIFICADO DE MÁXIMA UTILIDAD DE LA INFORMACIÓN

3.1. DEFINICIÓN. Dada una población finita de tamaño N, dividida en k estratos de tamaños $N_1, N_2,\ldots,N_k$, la cual se considera una fuente de información, de la cual se selecciona una muestra de tamaño n, distribuida en los estratos por submuestras de tamaño $n_1,n_2,\ldots,n_k$, respectivamente en cada estrato. Definimos la utilidad total de la información de la muestra de la siguiente manera:

$$U = \sum_{i=1}^{k} U_i \ln(p_i) \quad (3.1)$$

Donde $p_i = \frac{n_i}{n}$ y U_i es la utilidad de la información del estrato i.

Nota: La utilidad de la información del estrato i: U_i puede modelarse por H_i o pueden usarse modelos más complejos.

3.2. PROBLEMA DE LA MÁXIMA UTILIDAD DE LA INFORMACIÓN

Por tanto, el problema de la máxima información es:

$$\begin{aligned} &Max \;\; U = \sum_{i=1}^{k} U_i \ln(p_i) \\ &sujeto \; a: \\ &\sum_{i=1}^{k} p_i = 1 \\ &\forall i = 1,2,\ldots,k : p_i \geq 0 \end{aligned} \quad (3.2)$$

Este es claramente un problema de optimización no lineal, el cual tiene una solución bastante simple, usando las condiciones de Kuhn-Tucker. Para usar las condiciones, recordemos que un problema de programación no lineal en general puede formularse de la siguiente manera:

Encontrar un vector $x = (x_1, x_2, \ldots, x_n)$ para:

$$\begin{aligned} &Max \;\; f(x) \\ &sujeto \; a: \\ &h(x) = 0 \\ &x \geq 0 \end{aligned} \quad (3.3)$$

Teorema de Kuhn-Tucker. Sea $\mathbf{x}^*$ un extremo local de $f(\mathbf{x})$ que satisface las condiciones $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ y además es un punto regular de esas restricciones. Entonces, existe $\lambda \in R^m$ tal que:

$$\nabla f(x) + \lambda^T \nabla h(x^*) = 0 \quad (3.4).$$

Si aplicamos el teorema a nuestro problema, se tiene:

$$p^* = (p_1^*, p_2^*, \ldots, p_n^*)$$

$$\nabla f(p^*) = \begin{pmatrix} \frac{U_1}{p_1^*} \\ \frac{U_2}{p_2^*} \\ . \\ . \\ . \\ \frac{U_n}{p_n^*} \end{pmatrix} \qquad \nabla h(p^*) = \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$$

De donde (3.4) se escribe así:

$$\begin{aligned} &\frac{U_1}{p_1^*} + \lambda = 0 \\ &\frac{U_2}{p_2^*} + \lambda = 0 \\ &. \\ &. \\ &. \\ &\frac{U_n}{p_n^*} + \lambda = 0 \end{aligned} \qquad (3.5)$$

Este sistema puede convertirse en:

$$\begin{aligned} &U_i = -\lambda p_1^* \\ &U_2 = -\lambda p_2^* \\ &. \\ &. \\ &. \\ &U_n = -\lambda p_n^* \end{aligned} \qquad (3.6)$$

Sumando los dos lados del sistema, se tiene:

$$\sum_{i=1}^{n} U_i = -\lambda \sum_{i=1}^{n} p_i^*$$

de donde:

$$\lambda = \frac{-\sum_{i=1}^{n} U_i}{\sum_{i=1}^{n} p_i^*}; \text{ pero como } \sum_{i=1}^{n} p_i^* = 1; \text{ entonces:}$$

$$\lambda = -\sum_{i=1}^{n} U_i. \text{ Por tanto:}$$

$$p_1^* = \frac{U_1}{\sum_{i=1}^{n} U_i};\ p_2^* = \frac{U_2}{\sum_{i=1}^{n} U_i}; \ldots; p_n^* = \frac{U_n}{\sum_{i=1}^{n} U_i} \qquad (3.7)$$

Se puede verificar que esta solución cumple las condiciones de Kuhn- Tucker de segundo orden y por tanto es la óptima.

Condiciones de suficiencia de segundo orden. Supóngase que hay un punto x^* que satisface $h(x^*) = 0$ y un $\lambda \in R^n$ tal que:

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0 \quad (3.8)$$

Supóngase también que la matriz $L(x^*) = F(x^*) + \lambda^T H(x^*)$ es definida negativa en $M = \{y \mid \nabla h(x^*) y = 0\}$, esto es, para $\boldsymbol{y} \in \boldsymbol{M}$, $\boldsymbol{y} \neq \boldsymbol{0}$, se cumple que $\mathbf{y}^T \boldsymbol{L}(\mathbf{x}^*)\mathbf{y} < \mathbf{0}$. Entonces, x^* es un máximo local estricto del problema. Si aplicamos este teorema a nuestro caso tenemos:

$$\begin{pmatrix} \frac{U_1}{p_1^*} \\ \frac{U_2}{p_2^*} \\ . \\ . \\ . \\ \frac{U_n}{p_n^*} \end{pmatrix} - \sum_{i=1}^{n} U_i \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{pmatrix} = 0$$

y sabiendo que:

$$p_1^* = \frac{U_1}{\sum_{i=1}^{n} U_i};\ p_2^* = \frac{U_2}{\sum_{i=1}^{n} U_i}; \ldots; p_n^* = \frac{U_n}{\sum_{i=1}^{n} U_i}, \quad \text{se obtiene:}$$

$$\begin{pmatrix} \frac{\frac{U_1}{U_1}}{\sum_{i=1}^{n} U_i} \\ \frac{\frac{U_2}{U_2}}{\sum_{i=1}^{n} U_i} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\frac{U_n}{U_n}}{\sum_{i=1}^{n} U_i} \end{pmatrix} - \sum_{i=1}^{n} U_i \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix} = 0$$

$$\begin{pmatrix} \sum_{i=1}^{n} U_i \\ \sum_{i=1}^{n} U_i \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} U_i \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^{n} U_i \\ \sum_{i=1}^{n} U_i \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} U_i \end{pmatrix} = 0$$

Por otro lado,

$$F(p^*) = \left(\frac{\partial^2 f(p^*)}{\partial p_i \partial p_j} \right) = \begin{pmatrix} -U_1 / p_1^{*2} & & 0 \\ & \ddots & \\ 0 & & -U_n / p_n^{*2} \end{pmatrix}$$

Similarmente: $H(p^*) = \left(\dfrac{\partial^2 h(p^*)}{\partial p_1 \partial p_j} \right) = 0$

Luego: $L(p^*) = F(p^*)$

Además:

$$M = \{ y | \nabla h(x^*) y = 0 \} = \left\{ y \left| \sum_i y_i = 0 \right. \right\}$$

Por otro lado:

$$y^T L(p^*) y = (y_1 \dots y_n) \begin{pmatrix} -U_1 / p_1^{*2} & & 0 \\ & \ddots & \\ 0 & & -U_n / p_n^{*2} \end{pmatrix} \begin{pmatrix} y_1 \\ \\ y_n \end{pmatrix} < 0$$

Lo que significa que: $\sum_i \dfrac{-U_i}{p_i^{*2}} y_i^2 < 0$ si algún

$y_i^2 \neq 0$, lo cual se cumple cuando:

$$y \in M = \{ y | \nabla h(x^*) y = 0 \} = \left\{ y \left| \sum_i y_i = 0 \right. \right\}$$

y además **y** $\neq \boldsymbol{0}$.

4. CONCLUSIONES

En este artículo se ha demostrado que es posible otra metodología diferente a la del error muestral para el diseño del muestreo estratificado y es una basada en un modelo de la utilidad que a su vez se basa en la Teoría de la Información de Hartley y la entropía de Shannon. Otros enfoques son posibles y pueden conducir a soluciones diferentes, porque siendo la utilidad una función subjetiva puede ser modelada de muchas maneras incluyendo modelos empíricos que no han sido investigados y están fuera del alcance de este trabajo.

5. APENDICE

APLICACIÓN NUMÉRICA EN EXCEL

A continuación, se muestra un ejemplo numérico con una muestra de 400 sobre una población de 200000 con 10 estratos, donde se comparan la información total obtenida y el error muestral.

Se puede ver que la diferencia de la utilidad de información es de 42 [nats] pero la perdida de error muestral en el muestreo basado en la utilidad de la información es de apenas el 0.36%.

TABLA I

Diseño muestral estratificado basado en la utilidad de la información

Aplicación numérica en Excel

Estratos	fi	fi^2	ln(fi)	H	pi	ln(pi)	fi^2/pi	Muestra H	Muestra Prop.
A1	0.05	0.0025	-3.00	77.01	0.07	-2.66	0.04	28	20
A2	0.06	0.0036	-2.81	88.31	0.08	-2.53	0.05	32	24
A3	0.04	0.0016	-3.22	64.89	0.06	-2.84	0.03	23	16
A4	0.2	0.04	-1.61	197.16	0.18	-1.72	0.22	71	80
A5	0.3	0.09	-1.20	241.21	0.22	-1.52	0.41	87	120
A6	0.04	0.0016	-3.22	64.89	0.06	-2.84	0.03	23	16
A7	0.05	0.0025	-3.00	77.01	0.07	-2.66	0.04	28	20
A8	0.01	0.0001	-4.61	20.77	0.02	-3.97	0.01	8	4
A9	0.05	0.0025	-3.00	77.01	0.07	-2.66	0.04	28	20
A10	0.2	0.04	-1.61	197.16	0.18	-1.72	0.22	71	80
	1	0.1844	-27.27	1105.42	1.00	-25.13	1.07	400	400

TABLA II

Diseño muestral estratificado basado en la utilidad de la información

Ejemplo numérico

Tamaño Muestral	400
Tamaño Poblacional	200000
Información MaxInf.	-2336.15
Inform Prop.	-2379
Diferencia Información	42.8451
Error Muestral MaxInf	5.26%
Error Muestral Prop.	4.90%
Diferencia Error Muestral	0.36%

REFERENCIAS BIBLIOGRÁFICAS Y ELECTRÓNICAS

[1]. **LUENBERGER, D.** (1989). *"Programación lineal y no lineal"*. Addison-Wesley Iberoamericana, S. A., Wilmington, Delaware, Estados Unidos de América.

[2]. **EL-HAIK, B.** (2005). *"Axiomatic quality: integration axiomatic design with sixsigma, reliability and quality engineering"*. John Wiley and Sons Inc., Hoboken. New Jersey, Estados Unidos de América.

[3]. **SUH, N. P.** (1990). *"The Principles of Design"*. Oxford University Press, Oxford, Reino Unido.

[4]. **SUH, N. P.** (2005). *"Complexity Theory and Applications"*. Oxford University Press, Oxford, Reino Unido.

[5]. **COCHRAN, W** (1980). *"Técnicas de Muestreo"*. Compañía Editorial Continental, S. A. DE C. V., México, D. F., México