REVISTA DE ECONOMÍA Y ESTADÍSTICA
INSTITUTO DE ECONOMÍA Y FINANZAS
FACULTAD DE CIENCIAS ECONÓMICAS
UNIVERSIDAD NACIONAL DE CÓRDOBA
ISSN 0034-8066

ARTÍCULOS

Una Solución Alternativa al Problema de la Selección Controlada

Fernando Ferrero

Revista de Economía y Estadística, Tercera Época, Vol. 16, No. 1-2-3-4 (1972): 1º, 2º, 3º y 4º Trimestre, pp. 47-64.

http://revistas.unc.edu.ar/index.php/REyE/article/view/3683

UNA SOLUCION ALTERNATIVA AL PROBLEMA DE LA SELECCION CONTROLADA

FERNANDO FERRERO

I

El método de la selección controlada, notable creación de Goodman, R. y Kish, L., (1) constituye una herramienta de utilidad insustituible en el diseño de muestras por conglomerados. En las muestras de cobertura nacional, el Buró del Censo y el Survey Research Center de la Universidad de Michigan han utilizado profusamente esta técnica de selección, logrando hasta el presente resultados sumamente alentadores desde el punto de vista de un diseño óptimo.

Cuando se han agotado las posibilidades de estratificación —las que por otra parte están limitadas por el número de unidades primarias que se quieren seleccionar—, este método permite introducir nuevos controles, evitando de este modo la construcción de estratos adicionales, al tiempo que hace posible "controlar" el azar impidiendo o minimizando la aparición de combinaciones de unidades primarias indeseables.

Su carácter sustantivo radica en que permite ir un poco más allá del azar, habilitando al estadístico a asignar probabilidades arbitrarias a determinadas combinaciones de unidades primarias, esto es maximizando las combinaciones preferidas y minimizando las no preferidas —dentro de límites fijados por determinadas restricciones— y sin que por ello el carácter puramente aleatorio de la selección resulte alterado; la muestra así seleccionada sigue siendo aleatoria a pesar de que se hayan fijado probabilidades arbitrarias de antemano. Si bien es cierto que a primera vista tal afirmación puede aparecer simplemente absurda, el secreto radica en asignar probabilidades a las *combinaciones de unidades primarias* pero respetando al mismo tiempo las probabilidades que a estas últimas le hubieran correspondido si se hubiera usado un método de selección diferente.

La justificación práctica de este método se pone de manifiesto cuando de un total A de unidades primarias que componen el universo se extraen a unidades de primera etapa. En tal circunstancia el mayor número de estratos que se podrían formar sería exactamente a en cuyo caso cada estrato estaría representado por una unidad primaria (para la estimación de errores se consolidarán pares de estratos). Un número a de estrato representa pues la máxima estratificación posible, pero en la práctica acontece a menudo que una vez seleccionadas las unidades primarias —por lo general con probabilidades proporcionales a sus respectivos tamaños— la combinación resultante puede ser bastante desagradable desde el punto de vista de las variables no controladas explícitamente por el sistema de estratificación.

Para ilustrar con un ejemplo muy sencillo, considérese un universo formado por cuatro personas, dos varones y dos mujeres. Si se extraen muestras de tamaño dos, el espacio muestral quedará constituido por los siguientes puntos (combinaciones):

Muestra :	M_1M_2	M_1V_1	M_1V_2	M_2V_1	M_2V_2	V_1V_2
Prob. :	1/6	1/6	1/6	1/6	1/6	1/6

Supóngase además que a los fines del análisis una muestra compuesta por dos personas del mismo sexo representa una combinación no preferida y en consecuencia se quiere minimizar la probabilidad de un resultado de esta naturaleza. Para ello, el estadístico decide controlar el azar asignando una probabilidad igual a cero a la primera y última combinación, distribuyendo el resto de probabilidad entre las combinaciones con representantes de ambos sexos y seleccionando la muestra no del universo sino del conjunto de combinaciones que con un varón y una mujer se pueden formar.

Si se fija una probabilidad igual a 0,25 a las combinaciones restantes, puede comprobarse que la probabilidad de un individuo cualquiera sigue siendo 1/2, tal como debería ser en una muestra de dos individuos de un total de cuatro. En este caso particular, de las cuatro combinaciones se selecciona una al azar y con igual probabilidad. El resultado en un ejemplo como éste es sencillo y no requiere ningún método especial de cálculo.

Lamentablemente el problema se complica enormemente cuando las unidades primarias provienen de estratos previamente formados, cuando las probabilidades de selección son desiguales, cuando los controles que se establecen son múltiples y cuando se tiene un número grande de unidades a seleccionar.

El método de cómputo, que se denomina construcción de los patrones de selección, (2) está basado en una rutina que demanda mucha habilidad de parte del estadístico. Procede por iteraciones sucesivas y recién hacia el final se sabe si se llegará a una solución o no, cuando los patrones seleccionados no alcanzan a agotar las probabilidades asignadas o no cumplen con algunos de los controles impuestos. En caso negativo, por lo general hay que rehacer los cálculos y empezar todo de nuevo tantas veces como sea necesario para arribar a una solución final.

Por lo que respecta a su adaptabilidad a las computadoras, el procedimiento requiere una cantidad excesiva de criterios, en cada etapa plantea diferentes alternativas, todo lo cual hace que hasta el presente el problema no haya podido ser resuelto satisfactoriamente por medio de un programa de cómputo.

El propósito del presente trabajo es mostrar que el problema de la selección controlada puede ser traducido a un sistema lineal, identificando cada combinación con una variable diferente y resuelto por el método de la programación lineal, con ayuda de la técnica de la base artificial. Este método dispone, como es sabido, de programas de computación especialmente adaptados a tales fines.

Básicamente lo que hay que hacer es asignar una variable diferente a cada combinación y ensamblar estas variables dentro de una función de costo, cada una con un coeficiente distinto según sea el orden de prioridad de la combinación respectiva. Las combinaciones de alta prioridad aparecerán con coeficientes de costo muy bajos, los cuales naturalmente crecerán a medida que descienda el orden de preferencia de la combinación.

El objeto del programa será entonces minimizar una función lineal de costo sujeta a ciertas restricciones. Un tipo de restricciones serán aquellas que aseguren que la probabilidad de selección de cada unidad primaria será exactamente igual a la que le correspondería en caso de un muestreo simple de unidades primarias. A esta restricción se llega por la vía de sumar las probabilidades asignadas a las combinaciones que contienen una determinada unidad primaria, de tal suerte que habrá tantas restricciones como unidades primarias en el universo. O bien, para evitar ecuaciones redundantes se podrá prescindir de una unidad primaria por estrato, pero teniendo la precaución de incluir al final una ecuación que asegure que la suma de las probabilidades de todas las combinaciones sea igual a uno.

Por su parte, cada uno de los controles que se establezcan darán lugar a tantas restricciones como niveles contengan.

II

Supongamos que el total N de unidades primarias se distribuye en H estratos con un número N_h en h-ésimo estrato. Las unidades primarias habrán de recibir probabilidades desiguales y proporcionales a sus respectivos tamaños. Supongamos además que se agregan dos controles, cada uno con k_1 y k_2 niveles respectivamente.

Conforme al procedimiento que se describirá las unidades se identificarán por la posición de la celda en que se asignan y por el estrato a que pertenecen. Conviene poner de relieve que un tipo de control podría ser localización geográfica, en cuyo caso toda la región se subdivide en k_1 regiones o zonas; otro control podría ser el grado de urbanización de la unidad, alto, mediano o bajo, en tal caso $k_2 = 3$. Estos controles, claro está, se usarán si es que no han sido ya incorporados en el proceso previo de estratificación.

Supondremos que en cada celda o no hay ninguna unidad o a lo sumo una. Este supuesto no menoscaba la generalidad del planteamiento ya que en aquellos casos en que hay más de una unidad, tal celda dará lugar a tantas combinaciones como unidades contenga; si por el contrario, una celda no contiene unidad alguna, toda combinación que contenga esa celda recibirá una probabilidad nula. De todos modos el supuesto mencionado es al solo efecto de simplificar la presentación del problema pero en modo alguno afecta su generalidad.

Para el estrato h, $h = 1, 2, \ldots H.$, tendremos los datos que aparecen en la tabla siguiente. Tomando una unidad primaria por estrato se podrán formar con tales datos diferentes combinaciones,

$$A_{1r_{11}r_{12}} \cdot A_{2r_{21}r_{22}} \cdot A_{3r_{31}r_{32}} \ldots A_{hr_{h1}r_{h2}} \ldots A_{Hr_{H1}r_{H2}},$$

donde,

$$r_{h1} = 1, 2, \ldots k_{h1} \quad , \quad r_{h2} = 1, 2, \ldots k_{h2} \quad , \quad h = 1, 2, \ldots H$$

(Las magnitudes $m_{hr_{h1}r_{h2}}$ representan los tamaños relativos de las unidades primarias ($A_{hr_{h1}r_{h2}}$) o sus probabilidades de selección dentro del h-ésimo estrato (probabilidades proporcionales al tamaño)).

Cada una de estas combinaciones deberá ser luego clasificada en aceptable o no aceptable según cumpla con los controles impuestos. Si de acuerdo a los controles toda combinación debe contener entre c y $c+1$ unidades procedentes del nivel r_{h2} del segundo control, aquellas que tengan menos de c o más de $c+1$ unidades se clasificarán como inaceptables y en consecuencia serán eliminadas.

Una vez descartadas las combinaciones no aceptables, tal como se describirá más adelante, las restantes se clasificarán en diversas categorías conforme al grado de prioridad que merezcan:

A) Las que tienen unidades primarias procedentes de distintos controles (filas y columnas). Es decir que para todas las u.p. en la combinación se verifica:

Estrato <u>h</u>

Control 2 / Contr. 1	1	2	....	r_{h2}	....	k_{h2}	
1	A_{h11} m_{h11}	A_{h12} m_{h12}	....	$A_{h1r_{h2}}$ $m_{h1r_{h2}}$	....	$A_{h1k_{h2}}$ $m_{h1k_{h2}}$	$m_{h1.}$
2	A_{h21} m_{h21}	A_{h22} m_{h22}	....	$A_{h2r_{h2}}$ $m_{h2r_{h2}}$	....	$A_{h2k_{h2}}$ $m_{h2k_{h2}}$	$m_{h2.}$
$\vdots$			$\vdots$		$\vdots$		
r_{h1}	$A_{hr_{h1}1}$ $m_{hr_{h1}1}$	$A_{hr_{h1}2}$ $m_{hr_{h1}2}$	....	$A_{hr_{h1}r_{h2}}$ $m_{hr_{h1}r_{h2}}$	....	$A_{hr_{h1}k_{h2}}$ $m_{hr_{h1}k_{h2}}$	$m_{hr_{h1}.}$
$\vdots$			$\vdots$		$\vdots$		
k_{h1}	$A_{hk_{h1}1}$ $m_{hk_{h1}1}$	$A_{hk_{h1}2}$ $m_{hk_{h1}2}$	....	$A_{hk_{h1}r_{h2}}$ $m_{hk_{h1}r_{h2}}$	....	$A_{hk_{h1}k_{h2}}$ $m_{hk_{h1}k_{h2}}$	$m_{hk_{h1}.}$
	$m_{h.1}$	$m_{h.2}$	$\vdots$	$m_{h.r_{h2}}$	$\vdots$	$m_{h.k_{h2}}$	$m_{h..}$

$$r_{11} \neq r_{21} \neq r_{31} \ldots \neq r_{H1} \quad \text{y}$$
$$r_{12} \neq r_{22} \neq r_{32} \ldots \neq r_{H2}$$

Designaremos a cada una de estas combinaciones con las variables, y_1, y_2, ... y_y

B) Las que tienen a lo más una fila o una columna repetida; es decir:

$r_{h1} = r_{h+s,1}$ para un cierto s y las demás todas diferentes, o bien,
$r_{h2} = r_{h+s,2}$ para un cierto s y las demás todas diferentes.

A estas combinaciones las identificamos con

$$x_1, x_2, \ldots\ldots x_x$$

C) Las que tienen hasta tres filas o columnas repetidas,

$r_{h1} = r_{h+s,1} = r_{h+s+t,1}$ para un determinado (s, t) y todas las demás diferentes; o bien,

$r_{h2} = r_{h+s+t,2} = r_{h+s,2}$ para un determinado (s, t) y todas las demás diferentes. Designaremos a éstas por $z_1, z_2, \ldots\ldots z_z$.

Se podrían haber considerado también una fila y una columna repetida, o las que tienen dos filas y una columna, etc. No obstante, quien tenga a su cargo la selección deberá decidir qué ordenación de subíndices constituyen combinaciones diferentes y cuál será el rango que a cada una le corresponda. No debe olvidarse que puede haber tres o más controles en vez de dos en cuyo caso la decisión acerca de cuáles constituyen combinaciones prioritarias dependerá del tipo de investigación y de los datos con que se cuente.

De todos modos admitiremos que las combinaciones, por lo que respecta al grado de preferencia, pueden ordenarse en diversas categorías:

1a. prior. : $(y_1, y_2 \ldots\ldots y_y)$
2a. prior. : $(x_1, x_2 \ldots\ldots x_x)$
3a. prior. : $(z_1, z_2 \ldots\ldots z_z)$
4a. prior. : $(w_1, w_2 \ldots\ldots w_w)$

Para formar las restricciones del problema, habrá que tener en cuenta que la suma de las probabilidades asignadas a las combinaciones que contienen la unidad primaria $A_{hr_{h1}r_{h2}}$ deberá ser igual a $m_{hr_{h1}r_{h2}}$, que es exactamente la probabilidad que le corresponde a esta unidad dentro del h-ésimo estrato y tal condición deberá ser satisfecha para todo h, r_{h1} y r_{h2}. Sin embargo, para cada estrato se podrá excluir la restricción correspondiente a una determinada unidad primaria a condición de incluir una ecuación que asegure que la suma de las probabilidades asignadas a todas las combinaciones sea igual a uno.

Con estos requisitos se forman tantas restricciones como unidades primarias existan en el universo (con la sola excepción de una por estrato). A primera vista la tarea puede parecer formidable pero prácticamente las muestras por conglomerados tienen un número de unidades primarias por estrato que puede ser bastante manejable en la realidad. Incluso se pueden formar subgrupos de 5 ó 10 estratos y resolver el problema aisladamente en cada grupo así constituido. Si la muestra consistiera de 100 estratos, este procedimiento demandaría entre 20 y 10 problemas de programación lineal diferentes. Esta alternativa podría suplir los problemas derivados de la pérdida de la independencia entre los estratos o entre las selecciones de cada estrato.

De la misma forma que se establecieron restricciones para las unidades primarias, cabe hacer lo mismo para los totales marginales de los controles dentro de estratos.

Por lo general el número de niveles de cada control es igual en todos los estratos; vg., la misma división en zonas que se hace en el estrato h, se lleva a cabo también en el $h+1$, $h+2$, etc., o sea que desde este punto de vista $k_{h1} = k_1$ y $k_{h2} = k_2$, para $h = 1, 2 \ldots . H$.

Si en total existen $k_1 + k_2$ niveles en ambos controles, un número igual de restricciones se agregarán al problema. La suma de los totales marginales correspondiente al nivel j-ésimo del segundo control, por ejemplo,

$$m_{.j} = \sum_h m_{h.j} \quad , \quad j = 1, 2, \ldots . k_2$$

y el mayor número entero en él contenido, $m'_{.j}$, tal que $m_{.j} = m'_{.j} + d_{.j}$ $(0 \leqslant d_{.j} \leqslant 1)$, indicarán que toda combinación aceptable deberá contener $m'_{.j}$ o $m'_{.j} + 1$ unidades primarias provenientes del j-ésimo nivel del control 2, excluyéndose todas aquellas que no satisfagan tal requisito.

Los controles marginales se cumplen si la suma de todas las combinaciones que contienen exactamente $m'_{.j}$ u. p. del nivel j, es igual a $1\text{-}d_{.j}$, las que tienen $m'_{.j} + 1$ —o sean las restantes— constituyen el complemento de lo anterior y su suma en consecuencia deberá ser igual a $d_{.j}$. De las dos restricciones, sólo hace falta una, dado que ya existe una restricción que asegura que la suma de las probabilidades asignadas a todas las combinaciones es igual a uno. De este modo se establecen las ecuaciones restantes, tantas como $k_1 + k_2$ niveles existan, con lo cual se tiene la matriz de restricciones del problema. Sea $y = (y_1, y_2 \ldots y_y)$

$$x = (x_1, x_2 \ldots x_x)$$

$$. \quad \ldots\ldots\ldots\ldots$$

luego,

$$f_{111}(y, x, z, \ldots\ldots) = m_{1\,11}$$

$$f_{121}(y, x, z, \ldots\ldots) = m_{1\,21}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \qquad \ldots\ldots$$

$$f_{hr_{h1}r_{h2}}(y, x, z, \ldots) = m_{hr_{h1}r_{h2}}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \qquad \ldots\ldots$$

$$f_{Hk_1(k_2-1)}(y, x, z, \ldots) = m_{Hk_1(k_2-1)}$$

$$f_{d_{.1}}(y, x, z, \ldots\ldots) = d_{.1}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \qquad \ldots\ldots$$

$$f_{d_{.r_{h2}}}(y, x, z, \ldots) = d_{.r_{h2}}$$

$$f_{d_{r_{h1}}}.(y, x, z, \ldots) = d_{r_{h1}}.$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \qquad \ldots\ldots$$

$$f_{d_{.k_2}}(y, x, z, \ldots) = d_{.k_2}$$

$$G(y, x, z, \ldots\ldots) = 1$$

$$0 \leqslant y, \; 0 \leqslant x, \; 0 \leqslant z \; \ldots\ldots\ldots$$

Las funciones serán todas lineales y sus coeficientes serán unos o ceros dependiendo ello de la ecuación que se trate.

Además hay que definir la función objetivo,

$$\varnothing = c_1 (y_1 + y_2 \ldots + y_y) + c_2 (x_1 + x_2 \ldots + x_x) +$$
$$+ c_3 (z_1 + z_2 \ldots z_z) + c_4 (w_1 + w_2 \ldots w_w) \ldots\ldots$$

donde lógicamente los coeficientes $c = (c_1, c_2, \ldots\ldots)$, tales que $c_1 < c_2 < c_3 < c_4 \ldots$, se fijan de manera que en la solución óptima, $y = (y_1, y_2 \ldots\ldots y_y)$ aparezca con el mayor valor posible, luego $x = (x_1, x_2, \ldots x_x)$, y así sucesivamente con las restantes combinaciones hasta llegar a las de última preferencia, a las que se tratará de hacerlas ingresar con el menor valor posible.

III

Para ilustrar mejor la aplicación del método, desarrollaré un ejemplo ficticio en el cual —a pesar de su tamaño reducido— se incluyen todos los aspectos que se pueden presentar en un problema real.

Supóngase 10 unidades primarias de diferentes tamaños, 3 estratos, 2 controles, uno con dos niveles y el otro con tres y de ellos hay que extraer una muestra de tres unidades primarias, una por cada estrato, y con probabilidades proporcionales a sus respectivos tamaños (ver tabla 2).

La suma de los controles marginales, es

$$m_{.1} = 0.80$$
$$m_{.2} = 2.20$$
$$m_{A.} = 1.20$$
$$m_{B.} = 1.50$$
$$m_{C.} = 0.30$$

y significa que toda combinación aceptable tendrá 0 ó 1 u.p. de la columna primera ($m_{.1} = 0.80$); 2 ó 3 u.p. de la segunda columna $m_{.2} = 2.20$); 1 ó 2 de la 1ª, 4ª ó 7ª fila; 1 ó 2 de la 2ª, 5ª u 8ª fila, y 0 ó 1 u.p. de la 3ª, 6ª o 9ª fila. Significan también que la suma de las probabilidades asignadas a las combinaciones que contienen una u.p. de la columna 1 deberá ser igual a 0.80; la suma de las combinaciones que contienen dos u.p. de la 2ª columna deberá ser 0.80; la suma de las combinaciones que contienen una unidad primaria de las filas 1ª, 4ª ó 7ª, será 0.80, etc.

TABLA 2

Estrato	Control 2 / Control 1	1	2	$m_{i.}$
I	A	A_{111} $m_{111} = 0.20$	A_{112} $m_{112} = 0.20$	0.40
	B	A_{121} $m_{121} = 0.20$	A_{122} $m_{122} = 0.40$	0.60
	C	A_{131} $m_{131} = 0$	A_{132} $m_{132} = 0$	0.00
II	A	A_{211} $m_{211} = 0.10$	A_{212} $m_{212} = 0.70$	0.80
	B	A_{221} $m_{221} = 0.00$	A_{222} $m_{222} = 0.00$	0.00
	C	A_{231} $m_{231} = 0.00$	A_{232} $m_{232} = 0.20$	0.20
III	A	A_{311} $m_{311} = 0.00$	A_{312} $m_{312} = 0.00$	0.00
	B	A_{321} $m_{321} = 0.30$	A_{322} $m_{322} = 0.60$	0.90
	C	A_{331} $m_{331} = 0.00$	A_{332} $m_{332} = 0.10$	0.10
	$m_{.j}$	0.80	2.20	3.00

En este caso específico por medio de una tabla de triple entrada se pueden obtener las diferentes combinaciones (Tabla 3), asignando un valor cero a las combinaciones no aceptables, v. g., la combinación $A_{111}A_{211}A_{322}$ tiene valor cero porque de acuerdo a los controles, unidades primarias de la primera columna no puede haber más de una ($m_{.1} = 0.80$); igualmente, $A_{112}\,A_{232}\,A_{332}$ tiene el valor cero puesto que de las filas 3ª, 6ª y 9ª sólo pueden salir cero o una unidad, etc. De esta forma el número de combinaciones se reduce a 17, en tanto que las restantes han sido descartadas sea porque había un cero en la celda respectiva o porque eran simplemente inaceptables.

El problema inmediato es pues asignar prioridades a las distintas combinaciones. Para ello se considerarán los siguientes órdenes:

1) las combinaciones donde, $r_{11} \neq r_{21} \neq r_{31}$ y a lo más dos $r_{h2} = r_{h+s,2}$, $h = 1,2,3$, $s \neq 0$, se identificaron por (y_1, y_2, y_3, y_4).

2) las combinaciones donde, $r_{11} \neq r_{21} \neq r_{31}$ pudiendo haber hasta tres $r_{12} = r_{22} = r_{32}$, se designaron por (x_1, x_2).

3) las combinaciones en las que a lo más dos, $r_{h1} = r_{h+s\,1}$ y $r_{h2} = r_{h+t\,2}$ para $h = 1,2,3$ y $s,t \neq 0$, se designaron por $(z_1,z_2,z_3,z_4,z_5,z_6,z_7,z_8)$.

TABLA 3

			Estrato III		
Estr.		Estrato II	A_{321} (30)	A_{322} (60)	A_{332} (10)
I	A_{111} (20)	A_{211} (10)	---	---	---
		A_{213} (70)	---	z_1	---
		A_{232} (20)	---	y_1	---
	A_{121} (20)	A_{211} (10)	---	---	---
		A_{212} (70)	---	z_2	y_2
		A_{232} (20)	---	z_3	---
	A_{112} (20)	A_{211} (10)	---	z_4	---
		A_{212} (70)	z_5	w_1	---
		A_{232} (20)	y_3	x_1	---
	A_{122} (40)	A_{211} (10)	---	z_6	y_4
		A_{212} (70)	z_7	w_2	x_2
		A_{232} (20)	z_8	w_3	---

4) las que tenían tres $r_{12} = r_{22} = r_{32}$, sin interesar el primer contrcl, se las identificó por (w_1, w_2, w_3).

De este modo se estableció un orden decreciente de preferencia de y a w, y en consecuencia el problema se reducirá a encontrar cualquier solución que haga máximo el valor de y y mínimo el valor de x y repartiendo el resto entre x y z de forma que x reciba también el máximo compatible con las restricciones del problema.

Si no se emplea el método de selección controlada, las probabilidades de selección en caso de muestreo simple al azar de unidades primarias, serían las que aparecen en la Tabla 4. Una combinación cualquiera del tipo y, tendría una probabilidad de 0.054; una combinación cualquiera del tipo w, tendría una prob. de 0.300. Con la selección controlada se procurará que las primeras reciban una probabilidad tan grande y las últimas tan pequeña como se pueda.

TABLA 4

Comb.	Prob.	(1)
y_1	0.024	
y_2	0.014	
y_3	0.012	
y_4	0.004	0.054
x_1	0.024	
x_2	0.028	0.052
z_1	0.084	
z_2	0.084	
z_3	0.024	
z_4	0.012	
z_5	0.042	
z_6	0.024	
z_7	0.084	
z_8	0.024	0.378
w_1	0.084	
w_2	0.168	
w_3	0.048	0.300

En términos de programación lineal, el problema consistirá en minimizar la función $\varnothing$,

$$\varnothing = y_1 + y_2 + y_3 + y_4 + 10\,(x_1 + x_2) + 100\,(z_1 + z_2 + z_3 + z_4 + z_5 + z_6 + z_7 + z_8) + 1000\,(w_1 + w_2 + w_3)$$

sujeta a las siguientes restricciones, $0 \leqslant y, x, z, w$, y

$$\begin{array}{r} 1) \\ 2) \\ 3) \\ 4) \\ 5) \\ 6) \\ 7) \\ 8) \\ 9) \\ 10) \\ 11) \\ 12) \end{array}
\begin{bmatrix}
1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&1&0&0&0&0&1&1&0&0&0&1&0&0\\
0&1&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&1&0&1&0&0&0&0&0\\
0&1&0&0&0&1&1&1&0&0&1&0&1&0&1&1&0\\
0&0&1&0&0&0&0&0&0&0&1&0&1&1&0&0&0\\
1&0&0&0&1&0&1&1&1&1&0&1&0&0&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&0&1&0&0&0&1&1&0&0&1&0\\
1&1&1&1&1&1&1&0&0&1&1&0&0&0&1&0&0\\
0&0&0&0&0&0&1&1&0&1&1&1&1&0&1&1&0\\
0&0&0&0&1&1&0&0&0&0&0&0&0&0&1&1&1
\end{bmatrix}
\begin{bmatrix} y_1\\ y_2\\ y_3\\ y_4\\ x_1\\ x_2\\ z_1\\ z_2\\ z_3\\ z_4\\ z_5\\ z_6\\ z_7\\ z_8\\ w_1\\ w_2\\ w_3 \end{bmatrix}
=
\begin{bmatrix} 0.20\\ 0.20\\ 0.20\\ 0.10\\ 0.70\\ 0.30\\ 0.60\\ 1.00\\ 0.80\\ 0.50\\ 0.70\\ 0.20 \end{bmatrix}$$

(1) En realidad, las combinaciones z_3, z_8 y w_3, no cumplen con los requisitos de los controles y en consecuencia debieron ser eliminadas.

Es evidente que de la confrontación de las ecs. 11 y 12 que la suma de los valores de z tendrá que ser cuanto más 0.70 y exactamente 0.70 si la suma de las w es cero. En este caso la suma de las x e y no podrá ser mayor de 0.30. Las ecuaciones 11 y 12 son las más restrictivas por lo que respecta a los valores de x e y.

Puesto que las restricciones son ecuaciones, el problema de encontrar una solución básica factible se simplifica por medio de la introducción de variables artificiales en un número igual a la cantidad de ecuaciones de la matriz. Sin embargo, para obviar la introducción de 12 variables adicionales, opté por identificar seis vectores bases (aplicando operaciones elementales a la matriz original) y con ellos sólo fue necesario introducir seis variables artificiales. Conviene destacar que a la misma solución se llegaría continuando la elección de vectores bases —o sea reduciendo la matriz a su forma Hermítica— en cuyo caso si el rango de la matriz es doce, no serán necesarios introducir los vectores artificiales. La tabla inicial del Simplex tendrá de esta forma 17 variables, una por cada combinación, y seis variables artificiales.

La solución que se obtiene —degenerada, desde luego—, es

Comb.:	y_1	y_2	y_3	y_4	x_1	x_2	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	w_1	w_2	w_3
Prob.:	0	0	0	10	20	0	20	20	0	0	0	0	30	0	0	0	0

que da un valor para la función objetivo igual a $7210 = \varnothing$. Una de las soluciones óptimas es pues,

	Pr.		Pr.
$A_{122}A_{211}A_{332}$	0.10	$A_{121}A_{212}A_{322}$	0.20
$A_{112}A_{232}A_{322}$	0.20	$A_{122}A_{212}A_{321}$	0.30
$A_{111}A_{212}A_{322}$	0.20		1.00

Si comparamos estas probabilidades con las de la tabla 4, se comprobará que las de más alta prioridad han aumentado de 0.054 a 0.10; las de 2ª prioridad, de 0.052 a 0.20. En conjunto, la suma de x e y se ha casi triplicado; a su vez, las de menor preferencia

			y_1	y_2	y_3	y_4	x_1	x_2	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	w_1	w_2	w_3	P_1	P_2	P_3	P_4	P_5	P_6	
			1	1	1	1	10	10	100	100	100	100	100	100	100	100	1000	1000	1000	u	u	u	u	u	u	
y_1	1	20	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
P_1	u	70	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	1	0	0	0	0	0	78
y_2	1	20	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
y_4	1	10	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	13
x_2	10	50	0	0	0	0	0	1	1	0	—1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	55
P_2	u	50	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	1	1	0	1	0	0	0	0	57
P_3	u	70	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	0	1	0	0	0	78
P_4	u	50	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	1	1	0	0	0	1	0	0	57
P_5	u	30	0	0	0	0	0	0	—1	1	0	—1	—1	1	1	0	—1	1	0	0	0	0	0	1	0	30
y_3	1	30	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	34
P_6	u	70	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	1	78
x_1	10	40	0	0	0	0	1	0	0	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0	48
		980	0	0	0	0	0	0		—89		—89		—89		—99		—980								—2702
									—89		—99		—89		—89		—980		—990							
		340	0	0	0	0	0	0	2	6	2	2	2	6	6	2	2	6	2	0	0	0	0	0	0	378

bajan de 0.30 a 0. Evidentemente el resto, 0.70, queda asignado a las combinaciones z, lo cuál era necesario de conformidad a las restricciones 11 y 12.

De igual forma se pueden obtener otras soluciones que hagan $w = 0$, y con diferentes alternativas para las restantes variables. Las siguientes han sido obtenidas por el mismo proceso y tienen el mismo carácter de óptimas:

Comb.	y_1	y_2	y_3	y_4	x_1	x_2	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	w_1	w_2	w_3
1ª:	0	0	0	10	20	0	20	20	0	0	0	0	30	0	0	0	0
2ª	0	10	0	0	20	0	20	10	0	0	0	10	30	0	0	0	0
3ª	10	0	0	0	10	10	10	20	0	0	10	10	20	0	0	0	0

Cualquier combinación de estas últimas es también una solución al problema. Así por ejemplo se puede obtener una solución con diez combinaciones no nulas (multiplicando la suma de las dos primeras por 1/4 y sumándole la tercera multiplicada por 1/2).

Comb.	Prob.	
$A_{111}A_{232}A_{322}$	0.050	
$A_{121}A_{212}A_{332}$	0.025	
$A_{122}A_{211}A_{332}$	0.025	
$A_{112}A_{232}A_{322}$	0.150	
$A_{122}A_{212}A_{332}$	0.050	
$A_{111}A_{212}A_{322}$	0.150	
$A_{121}A_{212}A_{322}$	0.175	
$A_{112}A_{212}A_{321}$	0.050	
$A_{122}A_{211}A_{322}$	0.075	
$A_{122}A_{212}A_{321}$	0.250	1.000

y esta solución dá para la función objetivo un valor $\varnothing = 7210$.

Si se selecciona una solución con valores no nulos para y_3 y z_4 y se la combina con la última se llegará a una solución no degenerada.

Eligiendo una combinación con probabilidades proporcionales al tamaño, fácil es comprobar que éstas deberán satisfacer todas las restricciones. La probabilidad de A_{111} es la suma de la 1ª y 6ª combinación, $0.05 + 0.15$, que es lo que le correspondería de acuerdo a los valores de la tabla 2. La suma de las probabilidades que tienen una unidad primaria en la columna 1 (1º nivel del 2º control) deberá ser 0.80; la suma de las probabilidades de las comb. 1, 2, 3, 5, 6, 7, 8, 9 y 10 es precisamente 0.80. Otro tanto se puede hacer con los restantes controles y comprobar si todos han sido satisfechos.

REFERENCIAS

(1) GOODMAN, R., y KISH, L., (1950), "Controlled Selection, a technique in probability sampling", *Journal of the American Statistical Assoc.*, vol. 45, pp. 350-372.

(2) HESS, I., (1963), *The Controlled Selection Technique,* Survey Research Center, Univ. of Michigan.

(3) HESS, I., y SRIKANTAN, K. S., (1966), *Some Aspects of the Probability Sampling Technique of Controlled Selection,* Survey Research Center, Univ. of Michigan.