

ESTIMACIÓN PERIÓDICA DEL CONSUMO DE UN SUMINISTRO PÚBLICO

M. G. Fernández, F. Guerrero y M. Gutiérrez Moya
Departamento de Organización Industrial y Gestión de Empresas
Escuela Superior de Ingenieros. Universidad de Sevilla
Avda. Reina Mercedes, s.n. 41012- Sevilla. E-mail: fergue@esic03.us.es

RESUMEN

En este trabajo se presenta el diseño e implementación de un plan de muestreo para la estimación periódica del consumo de un suministro público en un área metropolitana de una gran ciudad. Los tamaños muestrales en cada uno de los estratos según el criterio de afijación óptima. Además, se ha desarrollado un modelo de simulación del consumo de los suministros que permitió validar el procedimiento seguido antes de la explotación de datos reales tal como se está haciendo actualmente.

PALABRAS CLAVE: Estimación, muestreo estadístico.

1. INTRODUCCIÓN.

El problema que se aborda en este trabajo es la estimación periódica del consumo de un Suministro Público (SP) en un área metropolitana de una gran ciudad. La Empresa Suministradora (ES) abastece la capital y una serie de localidades circundantes. Los suministros se distribuyen en tres tipos de tarifas (A, B y C) en función del uso del mismo. La ES lee los contadores de los abonados y les factura cada tres meses.

En el caso de que la ES pretenda conocer el consumo correspondiente a un período inferior a los tres meses, tendrá que desarrollar algún procedimiento de estimación adecuado al problema. En particular, se ha abordado la estimación mensual del consumo de cada una de las tres tarifas, tanto para el área metropolitana, como para la capital de la misma.

El tratamiento elegido para estimar el consumo está basado en técnicas de muestreo estadístico, frente a otro posible enfoque basado en series temporales, el cual presenta dificultades en este consumo específico debido a la falta de estacionariedad de la serie y a las escasas posibilidades de lograr tal estacionariedad mediante las transformaciones habituales.

Los datos de partida para el diseño del plan de muestreo son los siguientes: para cada tarifa, la población de suministros aparece dividida en intervalos de consumo anual. Los intervalos de consumo anual tienen una amplitud de 25 unidades de consumo (uc) hasta llegar a los 1000 uc, y de 100 uc hasta los 20000 uc. El último intervalo abarca aquellos suministros con consumo anual mayor de 20000 uc. Para

cada uno de los intervalos de consumo se conoce el número de suministros, así como el consumo agregado anual de éstos.

Este trabajo está organizado de la siguiente forma. En la sección 2 se estudia la determinación de las fronteras de los estratos. En la sección 3 se presenta el procedimiento para el cálculo de los tamaños muestrales y la estimación del consumo. En la sección 4 se muestran diversos experimentos que permiten validar el procedimiento seguido a partir de la simulación del consumo de los suministros. Por último, en la sección 5 presentan las conclusiones obtenidas.

2. DETERMINACIÓN DE LAS FRONTERAS DE LOS ESTRATOS.

En numerosas aplicaciones del muestreo es frecuente recurrir a la estratificación a fin de reducir la varianza total y, por lo tanto, el error de muestreo mediante la utilización de información adicional que poseamos sobre el colectivo a muestrear. Como es sabido, las ventajas del muestreo estratificado son tanto mayores cuanto más homogéneas sean las unidades dentro de cada uno de los estratos y cuanto más heterogéneos sean estos últimos entre sí, lográndose una reducción de la varianza incluso en una afijación tan simple como la proporcional, la cual prescindiendo, de la corrección para poblaciones finitas, podemos aproximar mediante:

$$\text{Var}(\bar{x}_{ia}) - \text{Var}(\bar{x}_{ep}) \cong \frac{\sum_{h=1}^L n_h (\mu_h - \mu)^2}{n^2} \quad (2.1)$$

siendo \bar{x}_{ia} y \bar{x}_{ep} las respectivas medias muestrales de los diseños irrestrictamente aleatorio y estratificado con afijación proporcional; n_h y n son respectivamente los tamaños muestrales del estrato h -ésimo ($h=1, 2, \dots, L$) y del total de la muestra; μ_h y μ designan, respectivamente, las medias poblacionales del estrato h -ésimo y de la población global.

Supuestos definidos los estratos, la teoría del muestreo estratificado es muy conocida y puede encontrarse en cualquier manual de Estadística. El verdadero problema en nuestro caso radicaba en que la especificación de las fronteras de los estratos (cuestión que se supone definida a priori en los manuales) no tiene una solución analítica única, sino que goza de una gran flexibilidad debido a que es un proceso en gran medida subjetivo y, de hecho, interacciona con el procedimiento de afijación que se establezca.

Algunas de las directrices o sugerencias que tratan de auxiliar en esta ambigüedad no proporcionan gran ayuda. Excluido cualquier criterio cualitativo (geográfico, etc.) que hubiera podido significar un método "natural" de estratificación, la cota

superior, $\frac{n}{2}$, para el número de estratos (a fin de lograr una estimación insesgada de la varianza) representa un número excesivo que no sólo dificultaría la operatividad en la recogida de datos muestrales sino que, además, sólo lograría ganancias marginales insignificantes en cuanto a reducción del error muestral, ya que la porción R^2 de la varianza afectada por la estratificación disminuye con el número de estratos de acuerdo con el modelo $\frac{R^2}{L^2 + (1 - R^2)}$, siendo L el número de estratos, y en cambio la porción $(1 - R^2)$ de la varianza no relacionada con la variable de estratificación no se ve afectada al aumentar el número de estratos. Puesto que el estrato de mayor consumo anual evidencia una gran variabilidad se decidió muestrear al 100% (es decir, con certeza) este estrato superior, con lo cual es nula su contribución a la varianza global. Por otra parte, la interacción mencionada anteriormente está presente en las dos sugerencias clásicas sobre la especificación de fronteras para los estratos:

- a) La de Hansen, Hurwitz y Madow que aconsejan establecer la frontera del estrato de certeza aproximadamente en el punto $Y/2n$, siendo Y la cantidad total de la variable de interés.
- b) La de Delanius que, aún sin considerar el diferente coste en la toma de muestras de unidades correspondientes a distintos estratos, recomienda que la frontera óptima y_h entre los estratos h y $h+1$ satisfaga la relación:

$$\frac{(y_h - \bar{Y}_h)^2 + S_h^2}{S_h} = \frac{(y_h - \bar{Y}_{h+1}) + S_{h+1}^2}{S_{h+1}} \quad (2.2)$$

Obviamente esta solución "óptima" no resulta práctica porque los cinco parámetros de la ecuación dependen de y_h cuyo valor es desconocido.

A fin de "objetivar", en lo posible, un problema que de por sí es abierto, hemos recurrido a un procedimiento de "feed-back" lográndose así, al aplicar la afijación óptima de Neyman, una representación "equilibrada" de los distintos tipos del universo de consumidores objeto de interés.

3. CÁLCULO DE TAMAÑOS MUESTRALES Y ESTIMACIÓN DEL CONSUMO.

Una vez fijados los estratos para cada una de las tarifas, el siguiente paso será determinar el tamaño muestral global para cada una de ellas. A continuación se presentan los fundamentos del cálculo del tamaño muestral para una tarifa genérica.

Tabla 1. Nomenclatura para el cálculo del tamaño muestral

Subíndices:	
h	intervalos
k	estratos
Definición de estratos:	
L	número total de estratos
L'	número del último estrato abarcado por el muestreo
Número de suministros del fichero histórico:	
NS _h	número de suministros del intervalo h
NS _k	número de suministros del estrato k
NS _T	total de suministros del fichero histórico
N'	total de suministros en los estratos sometidos a muestreo
Consumos del fichero histórico:	
C _h	consumo total de los suministros del intervalo h
C _k	consumo total de los suministros del estrato k
Consumos medios del fichero histórico:	
CM _h	consumo medio del intervalo h
CM _k	consumo medio del estrato k
Medida de la dispersión (o variabilidad) del consumo en cada estrato:	
S _k	desviación típica del consumo en el estrato k
Tamaño muestral:	
n _h	tamaño muestral en el intervalo h
n _k	tamaño muestral en el estrato k
n'	tamaño muestral de los estratos sometidos a muestreo
n	tamaño muestral total
Parámetros de la estimación:	
e	error absoluto
Z	parámetro cuyo valor viene fijado unívocamente dando el grado de confianza de la estimación

El apóstrofe " ' " aplicado a un símbolo, hace que este se refiera a los estratos abarcados por el muestreo. A partir del número de suministros de cada intervalo, el número de suministros en cada estrato k viene dado por:

$$NS_k = \sum_{\forall h \in k} NS_h \quad \forall k \quad (3.1)$$

De igual forma, el consumo total de los suministros de cada uno de los estratos vendrá dado por:

$$C_k = \sum_{\forall h \in k} C_h \quad \forall k \quad (3.2)$$

El consumo medio en cada uno de los intervalos, dados el número de suministros del intervalo y el consumo total de estos, viene dado por:

$$CM_h = \frac{C_h}{NS_h} \quad \forall h \quad (3.3)$$

De igual forma, el consumo medio en cada uno de los estratos se calcula según:

$$CM_k = \frac{C_k}{NS_k} \quad \forall k \quad (3.4)$$

La desviación típica del consumo para cada uno de los estratos viene dada por:

$$S_k = \sqrt{\sum_{h \in k} (CM_h - CM_k)^2 \frac{NS_h}{NS_k}} \quad \forall k \quad (3.5)$$

El número total de suministros sometidos a muestreo es igual a la suma del número de suministros de los estratos de muestreo (los L' primeros). Expresando lo anterior matemáticamente:

$$N' = \sum_{k=1}^{L'} NS_k = NS_T - \sum_{k=L'+1}^L NS_k \quad (3.6)$$

El tamaño muestral total tendrá dos componentes, el tamaño muestral de los estratos sometidos a muestreo y el de los estratos de certeza:

$$n = n' + \sum_{k=L'+1}^L n_k \quad (3.7)$$

Obviamente para los estratos de certeza se verifica:

$$n_k = NS_k \quad k = L' + 1, \dots, L \quad (3.8)$$

En particular, en este trabajo se han considerado tres grados de confianza a los que les corresponden los valores de Z mostrados en la Tabla 2.

Tabla 2. Grados de confianza

GRADO DE CONFIANZA	Z
90%	1.64
95%	2
99.7%	3

El tamaño muestral en los estratos sometidos al mismo se calcula según:

$$n' = \frac{\left(\sum_{k=1}^{L'} NS_k S_k \right)^2}{\frac{e^2}{Z^2} + \sum_{k=1}^{L'} NS_k S_k^2} \quad (3.9)$$

Teniendo en cuenta el criterio de afijación óptima, el tamaño muestral en cada estrato de muestreo se calcula teniendo en cuenta el número de suministros en el mismo y la desviación típica del estrato según:

$$n_k = n' \frac{NS_k S_k}{\sum_{k=1}^{L'} NS_k S_k} \quad (3.10)$$

Es obvio que el tamaño muestral en cada estrato de muestreo debe ser un número entero. Además tenemos la restricción obvia:

$$n' = \sum_{k=1}^{L'} n_k \quad (3.11)$$

Cuando de acuerdo con el algoritmo, no todos los n_k resultan un número entero, se redondean con la restricción (3.11). Una vez hallados los n_k , se calculan los tamaños muestrales en cada uno de los intervalos. El reparto del tamaño muestral de un estrato k entre sus intervalos, se rige por la siguiente expresión:

$$n_h = n_k \frac{NS_h}{\sum_{\forall h \in k} NS_h} \quad (3.12)$$

De igual forma que para los estratos, los n_h deben ser números enteros y, para un estrato k la suma de los tamaños muestrales de los intervalos pertenecientes al estrato debe ser igual a n_k . En este caso se aplica un procedimiento de ajuste similar al descrito antes para los estratos.

Una vez determinados los tamaños muestrales por estratos e intervalos, el siguiente paso consiste en estimar el consumo total a partir de las lecturas muestrales obtenidas por los lectores de la ES. Además se determinará el intervalo de confianza de la estimación del consumo total.

Adicionalmente a la nomenclatura descrita en la Tabla 1, se utilizará la siguiente:

Tabla 3. Nomenclatura para la estimación del consumo total

Consumos:	
CR_k	consumo de los suministros de la muestra que pertenecen al estrato k
\overline{CR}_k	consumo medio de los suministros de la muestra que pertenecen al estrato k
\hat{C}	estimación del consumo total
Desviación de la estimación del consumo:	
$\sigma_{\hat{C}}$	desviación típica de la estimación del consumo
$V_{\overline{C}}$	varianza de la estimación del consumo medio en los intervalos sometidos a muestreo
$\sigma_{\overline{C}}$	desviación típica de la estimación del consumo medio en los intervalos sometidos a muestreo

Dados los consumos de cada uno de los suministros de la muestra, el cálculo de CR_k no es más que la suma de los consumos muestrales de suministros pertenecientes al

estrato k. Si dividimos CR_k por el número de suministros de la muestra que hay en el estrato k tendremos \overline{CR}_k . La estimación del consumo total es:

$$\hat{C} = \sum_{k=1}^{L'} NS_k \times \overline{CR}_k + \sum_{k=L'+1}^L CR_k \quad (3.13)$$

El intervalo de confianza del consumo total es:

$$\hat{C} \pm Z \sigma_{\hat{C}} \quad (3.14)$$

El parámetro Z es el mismo que el utilizado en la Tabla 2. Para el cálculo del intervalo de confianza en este trabajo se ha utilizado el Z correspondiente a un 95% de confianza. La desviación típica de la estimación del consumo la obtendremos a partir de su varianza, obteniéndose ésta según:

$$V_{\hat{C}} = \left(\sum_{k=1}^{L'} NS_k \right)^2 V_{\overline{C}} \quad (3.15)$$

La varianza de la estimación del consumo medio en los estratos sometidos a muestreo se calcula a partir de su desviación típica, la cual viene dada por la expresión siguiente:

$$\sigma_{\overline{C}} = \sqrt{\frac{1}{N'^2} \left(\sum_{k=1}^{L'} NS_k S_k \right)^2 - \frac{1}{N'^2} \sum_{k=1}^{L'} NS_k S_k^2} \quad (3.16)$$

4. APLICACIÓN DEL PLAN DE MUESTREO.

En este apartado se muestran los resultados de aplicar el plan muestreo descrito anteriormente, a la estimación del consumo del área metropolitana y de la capital de la misma. El número total de suministros de la ES es de 123180, 22565 y 940 para las tarifas A, B y C respectivamente. A continuación se ilustra la evolución del tamaño muestral, por un lado frente al error relativo, y por otro, frente al grado de confianza.

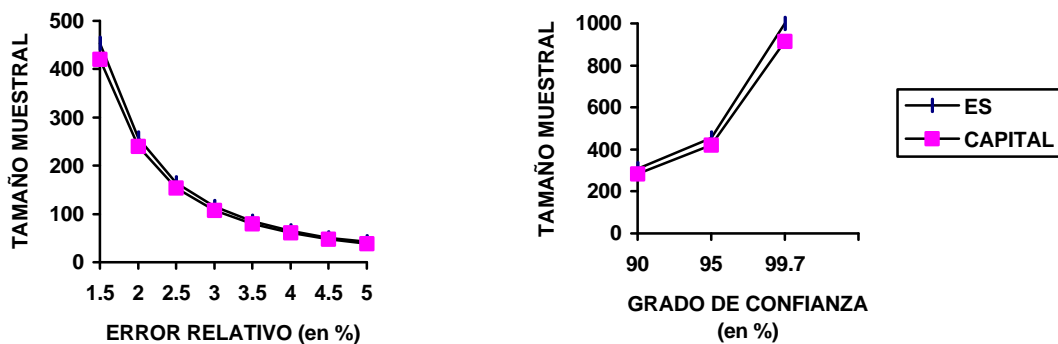


Figura 2. Variación del tamaño muestral

El diseño se verificó inicialmente mediante simulación dada la no disponibilidad de lecturas reales y el elevado coste de las mismas. Actualmente el modelo trabaja satisfactoriamente con datos reales.

Para cada ciclo de la simulación se presenta el consumo estimado (anual) en cada uno de los estratos de la tarifa A. La suma para todos los estratos nos da el consumo estimado (anual) para esa tarifa. Además, se presenta el consumo real del año base en cada estrato. Por último, se calculan, para cada estrato y para el total, las desviaciones (en porcentaje) del consumo estimado frente al real.

La nomenclatura utilizada en las tablas siguientes es:

- N.E.: número de estrato.
- C.R.: consumo real (en uc) del año base.
- C.S.1,...,C.S.5: estimación del consumo en cada una de los cinco ciclos de la simulación.
- D.S.1,...,D.S.5: desviación (en porcentaje) del consumo simulado frente al real en cada uno de los cinco ciclos de la simulación.
- TOTAL: agregación de los resultados (estimación y desviación) para todos los estratos.
- E.I. y E.F.: extremos inferior y superior del intervalo de confianza.

En la Tabla 4 se presentan los resultados de la simulación de la ES-Tarifa A. Observando las desviaciones se puede concluir que las estimaciones son precisas. Además las desviaciones referidas al total, no evidencian ninguna tendencia a sobrestimar o infraestimar, por lo que dado su carácter errático se puede asimismo concluir la no existencia de cualquier trayectoria que pudiera sugerir un sesgo sistemático.

Tabla 4. Simulación de la ES -Tarifa A

N.E.	C.R	C.S.1	C.S.2	C.S.3	C.S.4	C.S.5	D.S.1	D.S.2	D.S.3	D.S.4	D.S.5
1	1102525	1054116	1083397	1112678	1054116	1083397	-4,39	-1,73	0,92	-4,39	-1,73
2	4019434	4007206	4042668	3936282	4007206	4007206	-0,30	0,58	-2,07	-0,30	-0,30
3	5881530	5857728	5888237	5888237	5888237	5827219	-0,40	0,11	0,11	0,11	-0,92
4	4629327	4624472	4624472	4638571	4652670	4666769	-0,10	-0,10	0,20	0,50	0,81
5	3466008	3472716	3463902	3450681	3455088	3463902	0,19	-0,06	-0,44	-0,32	-0,06
6	5535899	5543457	5539626	5512809	5535795	5543457	0,14	0,07	-0,42	0,00	0,14
7	6810595	6851424	6854208	6843072	6843072	6840288	0,60	0,64	0,48	0,48	0,44
8	6007313	6012042	6005890	6016656	6010504	6008966	0,08	-0,02	0,16	0,05	0,03
9	6773103	6780688	6778632	6780688	6780688	6787884	0,11	0,08	0,11	0,11	0,22
10	2725056	2725056	2725056	2725056	2725056	2725056	0,00	0,00	0,00	0,00	0,00
TOTAL	46950790	46928905	47006088	46904730	46952432	46954144	-0,05	0,12	-0,10	0,00	0,01

En la Tabla 5, se presentan los extremos inferior y superior del intervalo de confianza para la estimación anual del consumo.

Tabla 5. Intervalo de confianza anual de la ES-Tarifa A

E.I.	46808476	46885057	46784377	46832240	46833823
E.F.	47049334	47127119	47025083	47072624	47074465

En la Tabla 6, se presentan los extremos inferior y superior del intervalo de confianza para la estimación mensual del consumo.

Tabla 6. Intervalo de confianza mensual de la ES-Tarifa A

E.I.	3900706	3907088	3898698	3902687	3902819
E.F.	3920778	3927260	3918757	3922719	3922872

Además, se han realizado diversos experimentos de simulación en los que se varían los parámetros de la misma. Todos los resultados validan el plan de muestreo diseñado.

5. CONCLUSIONES.

En este trabajo se aborda el problema de estimación del consumo de un suministro público en un área metropolitana de una gran ciudad. La herramienta utilizada para abordar el problema es el muestreo estadístico. Se ha diseñado el plan de muestreo, así como se ha validado mediante la utilización de un modelo de simulación. El modelo, en la práctica, ha demostrado su eficacia para inferir el consumo total al trabajar con datos reales.

6. REFERENCIAS.

- [1] **Jessen, R.J.** (1978), "Statistical Survey Techniques", Wiley .
- [2] **Deming, W.E.** (1970), "Sample Design in Business Research", Wiley.
- [3] **Kish, L.** (1972), "Muestreo de encuestas", Trillas.
- [4] **Mirás, J.** (1985), "Elementos de Muestreo para Poblaciones Finitas", Instituto Nacional de Estadística.
- [5] **Delanius, T. y Gurney, M.** (1951), "The problem of optimum stratification, II", Skandinavisk Aktuarietidskrift, 34, 133-148.