

# ***Diseño de Muestra Compleja para Estimación de Rating***

Luis A. Navarro H.; José B. Espinoza S. y Eduardo Quiroz

## **Resumen**

*Un diseño de muestra compleja se elabora para estimar el rating de audiencia de radio mediante una Encuesta panel aplicada cinco veces al año. El propósito de este diseño es la inclusión de variables de segmentación como sexo, edad y categoría familiar, lo que da origen a un muestreo por conglomerados de dos etapas y diferentes variables de estratificación. En la primera etapa se seleccionarán conglomerados de igual tamaño de 150 viviendas cada uno vía un muestreo estratificado, en la segunda etapa se seleccionaran dos viviendas por cada conglomerado mediante un muestreo aleatorio simple.*

## **1. INTRODUCCIÓN**

Los objetivos principales en la estimación del *rating* de audiencia de medios de comunicación son la determinación del diseño muestral y el cálculo de los errores muestrales. A continuación se presenta un esquema de diseño muestral para muestras complejas a ser utilizadas en el levantamiento de información. La sección 2 detalla el esquema del diseño muestral con las derivaciones de los pesos finales que van a utilizarse en los cálculos de las estimaciones de *rating* y errores muestrales; la sección 3 presenta resultados de la implementación de este diseño para un conjunto de datos piloto tomados de una empresa que elabora estos *rating*; la sección 4 presenta las conclusiones.

## **2. DISEÑO MUESTRAL**

La población piloto investigada es la que habita en viviendas familia-

res ubicadas en toda Lima Metropolitana. Se excluyen los que habitan en las llamadas viviendas colectivas, tales como orfanatos, asilos, cuarteles, cárceles, conventos, hospitales, comisarías, hoteles, etc. La frecuencia de aplicación de la Encuesta panel es de cinco veces por año. El Marco de la Encuesta se divide en *Marco de áreas*, constituido por las secciones distritales en que se encuentra dividido Lima Metropolitana, y *Marco de listas*, constituido por la relación de viviendas confeccionada para cada sección distrital seleccionada.

## **DEFINICIÓN DEL MUESTREO**

El muestreo es estratificado en dos grupos de niveles socioeconómicos, con el primer grupo o estrato conformado por los niveles socioeconómicos A y B, y el segundo grupo o estrato conformado por los niveles

socioeconómicos C, D y E. En cada estrato aplicamos un muestreo bietápico de conglomerados y estratificación de las unidades de la primera etapa.

### ESTRATIFICADO POR GRUPOS DE NIVELES SOCIO ECONÓMICOS

Recordemos que los estratos son grupos de población que se asumen homogéneos dentro de sí, pero posiblemente heterogéneos respecto a los otros estratos. La homogeneidad interna hace que no sea preciso extraer muestras muy grandes de cada uno de ellos, pero hay que preocuparse de que el diseño de estratos cubra todos los comportamientos diferentes que puedan preverse para los elementos de la población. Se debe destacar que este proceso de muestreo se aplica de modo *independiente* a cada estrato, permitiendo la aplicación simultánea de métodos de muestreo diferentes de acuerdo a la

información que se desea obtener.

### MUESTREO BIETÁPICO DE CONGLOMERADOS Y ESTRATIFICACIÓN EN LAS UNIDADES DE PRIMERA ETAPA.

Se trata de un muestreo bietápico porque se desarrolla en dos etapas. La primera se ejecuta en las secciones distritales que contendrán como unidades de selección a las UPM de 150 viviendas (conglomerado de 1ª etapa), asegurándose con ello que está representada toda la población de los distritos que conforman Lima Metropolitana y se garantiza la cobertura geográfica exigida; en la segunda etapa se hace un submuestreo de viviendas (conglomerados de 2ª etapa) de las UPM. Dentro de las unidades finales o viviendas de 2ª etapa se entrevista a todos los que aceptan participar.

Sobre las unidades de primera etapa o UPM se hace una estratificación de las secciones distritales que

1. Cercado Callao
2. Bellavista
3. Carmen de la Legua
4. La Perla
5. La Punta
6. Ventanilla



1. Cercado de Lima
3. Ate
4. Barranco
5. Breña
7. Comas
9. Chorrillos
10. El Agustino
11. Jesús María
12. La Molina
13. La Victoria
14. Lince
17. Magdalena del Mar
18. Miraflores
21. Pueblo Libre
22. Puente Piedra
25. Rímac
27. San Isidro
28. Independencia
29. San Juan de Miraflores
30. San Luis
31. San Martín de Porres
32. San Miguel
33. Santiago de Surco
34. Surquillo
35. Villa María del triunfo
36. San Juan de Lurigancho
38. Santa Rosa
39. Los Olivos
41. San Borja
42. Villa El Salvador
43. Santa Anita

permite su homogeneización y, por consiguiente, la mejora en la eficiencia del diseño.

### ESTRATIFICACIÓN DE LAS SECCIONES DISTRITALES

La estratificación se realiza atendiendo al objeto de la investigación y el criterio utilizado es el criterio geográfico que permite crear grupos homogéneos en cuanto a cobertura geográfica. Es decir, las secciones distritales serán los estratos en función al tipo de distrito, con unidades a seleccionar de grupos de 150 viviendas

### TAMAÑO DE LA MUESTRA

La calidad y la confiabilidad de los resultados de una encuesta por muestro reposan principalmente en el número de casos que se haya tomado como muestra. A este número se le conoce como *tamaño de la muestra*. Una muestra muy pequeña limita aquella confiabilidad, en tanto que una muestra muy grande eleva considerablemente los costos.

A fin de calcular el tamaño apropiado de la muestra para el presente estudio se consideró:

- El margen de error aceptado en las estimaciones: determinado estadísticamente para conocer el espacio cuantitativo dentro del cual se encuentran los valores reales del universo estudiado. Así, para el cálculo de la muestra total del presente estudio se ha determinado un margen de error de cuatro puntos porcentuales.
- El nivel de confianza: es el porcentaje de seguridad o probabilidad que se elige, que no es total,

pero sí se le considera como suficientemente razonable. Para el cálculo de la muestra del presente estudio se ha trabajado con un nivel de confianza del 95%.

- Tasa de no respuesta esperada: indica la pérdida muestral ocasionada por la inexistencia de respuesta. Así, para el cálculo de la muestra del presente estudio se ha trabajado con una tasa de no respuesta de 11%, lo cual arroja un coeficiente 1,11 para el ajuste del tamaño de la muestra.
- Efecto de diseño al usar muestreo con más de una etapa de selección: para el cálculo de la muestra del presente estudio se ha trabajado con un valor 1,12, estimado en base a los resultados de la Muestra panel en junio 2006.

$$n = [Z^2 p q N] / [d^2 (N - 1) + Z^2 p q] * efd * tnr$$

En teoría, el tamaño de la muestra es de 375 UPMs, lo que equivale a incorporar a la muestra 750 viviendas en los 39 distritos de Lima metropolitana.

### TAMAÑOS EN LA ESTRATIFICACIÓN

A cada estrato (cruce entre distrito y estrato social predominante) le corresponde un número de entrevistas proporcional al tamaño del número de hogares al que representa en su distrito. Para los distritos pertenecientes a los estratos AB, se ha realizado una afijación no proporcional con el fin de asegurar la incorporación de hogares de estratos altos.

TABLA 1: DISTRIBUCIÓN DE LA MUESTRA POR ESTRATO SOCIOECONÓMICO

NIVEL SOCIOECONÓMICO	MUESTRA DE VIVIENDAS	MUESTRA DE UPMs	% DE REPRESENTATIVIDAD DE LA UPM
ALTO/MEDIO ALTO	230	115	1,1
<b>AB</b>	<b>230</b>	<b>115</b>	<b>1,1</b>
BAJO	235	117	1,1
MUY BAJO MUY BAJO EXTREMO	285	143	1,3
<b>CDE</b>	<b>520</b>	<b>260</b>	<b>2,4</b>
<b>Total de la muestra</b>	<b>750</b>	<b>375</b>	<b>3,44</b>

(\*) Si en una vivienda hay más de un hogar, se selecciona el hogar principal.

### SELECCIÓN Y NÚMERO DE SECCIONES DISTRITALES

Se exige que cualquier vivienda a ser considerada en la muestra final tenga la misma probabilidad de ser seleccionada. Para ello, en la primera etapa las secciones distritales estarán constituidas por grupos de viviendas de tamaño 150 (UPM) y la selección de estas se hará de acuerdo a un muestreo estratificado. En la segunda etapa, se seleccionan dos viviendas de cada grupo de 150 viviendas seleccionadas previamente, mediante un muestreo aleatorio simple. Finalmente, se exigirá que las probabilidades finales de selección de cada vivienda sean iguales, lo que implica utilizar el criterio de probabilidad proporcional al tamaño o PPT (medido en número de viviendas) y así obtener muestras autoponderadas.

En la determinación de la estimación de los *rating* y errores muestrales correspondientes a las audiencias de las emisoras de radio, lectoría de diarios y revistas de Lima Metropolitana,

el criterio PPT determina los *pesos* que deben utilizarse para el cálculo de dichas estimaciones y errores muestrales según las exigencias de una muestra autoponderada. La forma de cálculo de dichos pesos se encuentra detallada a continuación.

### DETERMINACIÓN DE PESOS

El cuestionamiento más común de los clientes que compran información sobre audiencia, en la ejecución de una estimación de *rating* de audiencia de las emisoras de radio, lectoría de diarios y revistas de Lima Metropolitana, es sobre la igualdad de *chances* que tendrían las viviendas seleccionadas en la muestra final dado que esto conllevaría a posibles incrementos en el cálculo de los errores muestrales.

La respuesta a esto último es la utilización de un muestreo bietápico con selección PPT en la primera etapa que genera una muestra autoponderada con igual *chance* de selección en las unidades o viviendas finales.

Por lo tanto, la probabilidad de selección de la vivienda  $i$ , perteneciente a la UPM  $j$ , del estrato  $h$  al que se han afijado  $K_h$  UPM, y donde se ha determinado para la encuesta seleccionar un número fijo de  $K'_h$  viviendas por UPM, sería:

$$\begin{aligned}
 P(v_{ijh}) &= P(S_{jh})P(v_{ijh}|S_{jh}) \\
 &= K_h \frac{M_{jh}}{M_h} \times \frac{K'_h}{M_{jh}} \\
 &= K_h \frac{K'_h}{M_h} \times \frac{m_h}{M_h}
 \end{aligned}$$

Observe que el cálculo de esta probabilidad no depende de  $i$  ni de  $j$ . Además, se entiende que:

$P(S_{jh})$  : Probabilidad de selección de la UPM  $j$  en el estrato  $h$  por PPT

$P(v_{ijh} | S_{jh})$  : Probabilidad de selección de la vivienda  $i$  condicionada a la selección de la UPM  $j$ , en el estrato  $h$

Además,

$K_h$  : Número de UPM a tomar en el estrato  $h$

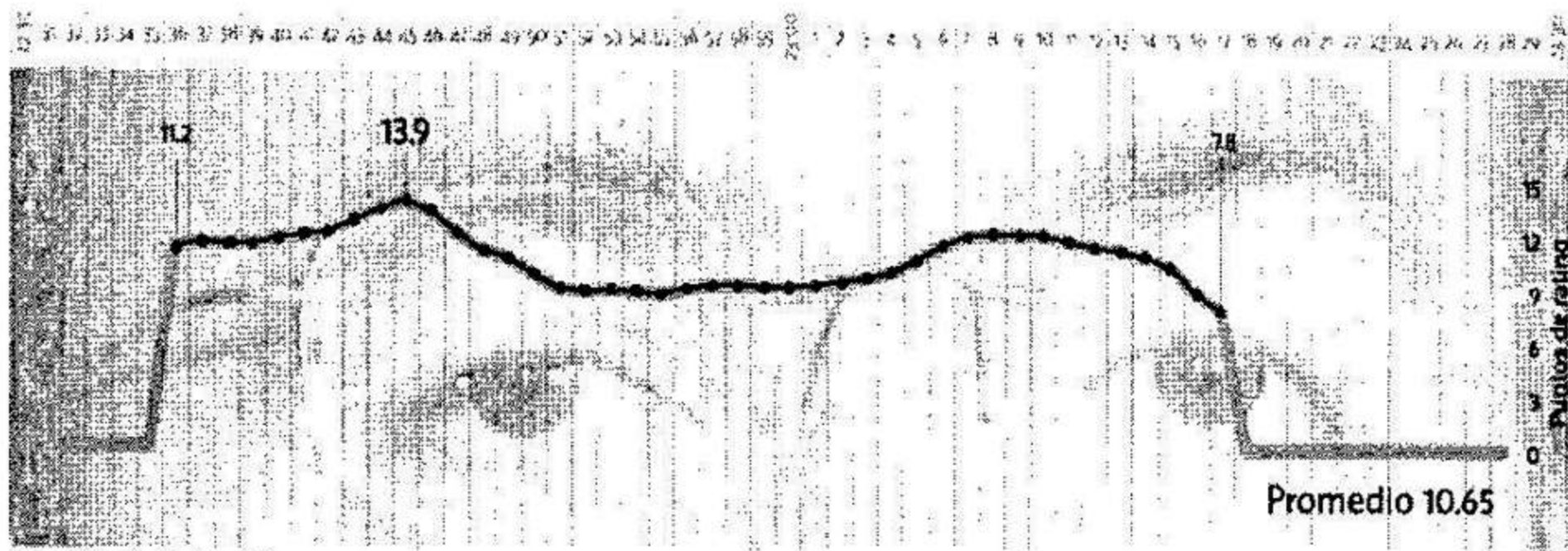
$K'_h$  : Número de viviendas seleccionadas (hogares a encuestar) por cada UPM en el estrato  $h$ .

$M_{jh}$  : Número total de viviendas en la UPM  $j$  del estrato  $h$ .

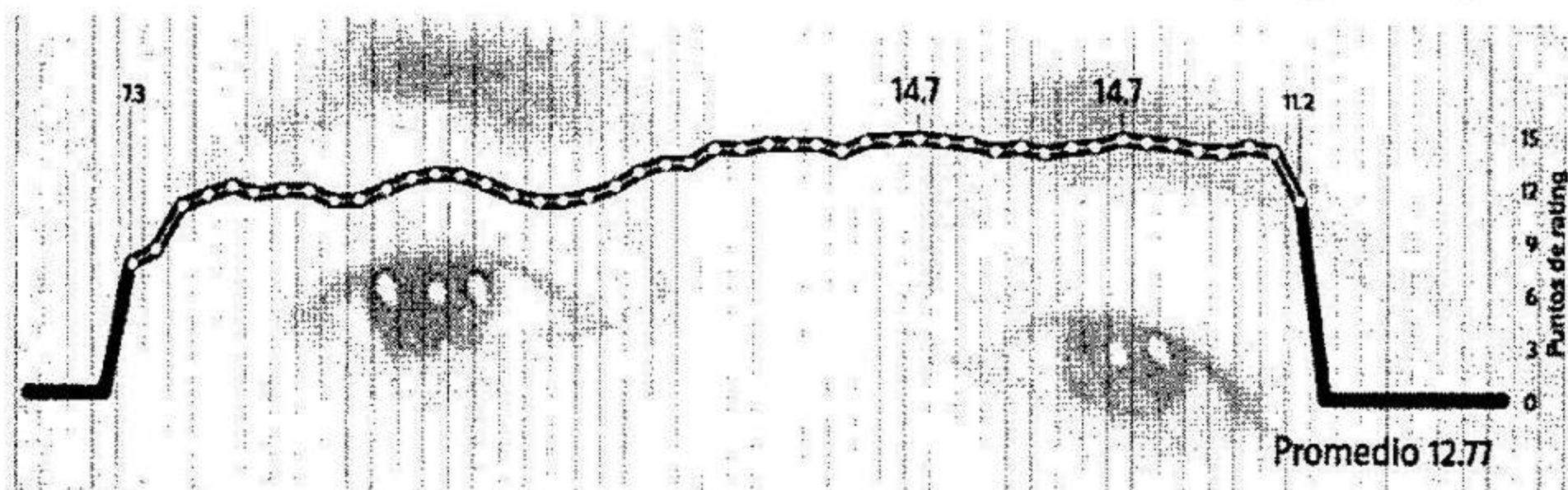
$M_h$  : Número total de viviendas en el estrato  $h$ .

$m_h = K_h K'_h$  : Número de viviendas (hogares a encuestar) en el estrato  $h$ .

Por ejemplo, el cálculo de la probabilidad de selección y peso final de



Felipe Calderón: Martes 14 de marzo



Roberto Madrazo: Martes 21 de marzo

una vivienda (de un total de dos viviendas finales en una UPM previamente seleccionada), en la tercera UPM (según las 3 UPM seleccionados vía fijación proporcional, como indica la tabla 1) en el distrito o estrato de Ate, para el nivel socioeconómico A y B, mediante un muestreo bietápico se

presenta en la tabla 2. Los cálculos de los pesos finales de viviendas pertenecientes a cada una de las UPM seleccionadas en los distritos correspondientes según la tabla 1, replicarán el número total de viviendas en cada uno de estos distritos.

**TABLA 2.- CÁLCULO DE PESOS FINALES POR VIVIENDA SELECCIONADA EN LA MUESTRA FINAL**

Estrato	$P(v_{ijh})$	Peso final de una vivienda por estrato	Número total de viviendas replicados por estratos
Ate	$3 \times \frac{150}{9115} \times \frac{2}{150}$	$\frac{9115}{3 \times 2}$	$\frac{9115}{3 \times 2} \times (3 \times 2) = 9155$
Barranco	$1 \times \frac{150}{3333} \times \frac{2}{150}$	$\frac{333}{1 \times 2}$	$\frac{3333}{1 \times 2} \times (1 \times 2) = 3333$
Bellavista	$2 \times \frac{150}{5094} \times \frac{2}{150}$	$\frac{5094}{2 \times 2}$	$\frac{5094}{2 \times 2} \times (2 \times 2) = 5094$
⋮	⋮	⋮	⋮

### 3. CONCLUSIONES

La contribución principal de este material consiste en la determinación de los pesos muestrales, uno de los cuestionamientos más comunes en los clientes que compran información sobre audiencia.

En la ejecución de una estimación de *rating* de audiencia de las

emisoras de radio, lectoría de diarios y revistas de Lima Metropolitana, el cuestionamiento se presenta sobre la igualdad de *chances* que tendrían las viviendas seleccionadas en la muestra final dado que esto conllevaría a posibles incrementos en el cálculo de los errores muestrales.

**BIBLIOGRAFÍA**

**NAVARRO, Luis** (2006) *Metodología para estimación de rating de audiencia*. Series NEC - UNI.

**JUDKIS, D. R.** (2000) *Fay's Method for variance Estimation*. Journal of official Statistics 6, 6:223-239,

**MARTINEZ ALMÉECIJA, A., J. MUÑOZ GARCÍA y A. PASCUAL ACOSTA.** (2004) *Tamaño de muestra y precisión estadística*. Universidad de Almería, Servicio de Publicaciones.

**Luis A. Navarro.** Doctor en Estadística. Consultor especializado en análisis cuantitativo. Docente UNI.

**José B. Espinoza.** Licenciado en Estadística. Consultor especializado en el manejo de diseños muestrales y experimentales, así como en análisis cuantitativo. Docente UNI

**Eduardo F. Quíroz.** Licenciado en Estadística. Consultor especializado en encuestas de opinión y mercadeo. Profesor de la FIECS - UNI y ex Director del Instituto de Investigaciones Económicas y Sociales.