

INVESTIGACIÓN DE MERCADOS: UN MODELO DE MUESTREO BIETAPICO ESTRATIFICADO PARA LA OBTENCIÓN DE ESTIMADORES EN VARIABLES CUALITATIVAS

Rafael A. Klinger Angarita

*Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali
rakling@pino.univalle.edu.co*

Resumen

El presente trabajo enfatiza sobre la importancia del diseño de una muestra probabilística para el desarrollo de una investigación de mercados, cuando las variables objeto de investigación son de tipo cualitativo. En particular se propone un método de muestreo bietápico estratificado para la estimación del parámetro proporción poblacional.

Palabras clave: Variable cualitativa, Parámetro, estimador, error de muestreo, nivel de confianza, selección aleatoria.

Abstract

The present work emphasizes on the importance of design of a probabilistic sampling method for marketing research, when the variables of investigation are of qualitative type. A two-stage stratified sampling method for the estimation of the population parameter is proposed.

Keywords: Qualitative variable, parameter, estimator, sampling error, confidence level, random selection.

1. Introducción

Los estudios de mercado toman por lo regular en consideración poblaciones objetivo muy grandes por lo que muchos investigadores las asumen como poblaciones infinitas, no porque en si lo sean desde el punto de vista matemático si no que al tomar una muestra de dicha población la fracción de muestreo (cociente entre el tamaño de muestra y el tamaño de la población) es muy inferior al 5%.

Todas las fases contempladas en una investigación de mercados son de suma importancia para la obtención de las conclusiones y la toma de decisiones frente al colectivo investigado; una de dichas fases es la concerniente a la definición del tamaño de la muestra y la respectiva selección de las unidades que serán encuestadas o medidas.

El presente artículo define un modelo de muestreo para la estimación de proporciones (indicador de variables cualitativas) que toma en cuenta los estratos socioeconómicos en que se divide cualquier población de consumidores y permite obtener indicadores estadísticos considerando dos unidades de muestreo (muestreo de conglomerados bietápico) (Pérez, 2005). Las unidades primarias de muestreo (conglomerados) que en el caso de una ciudad serán los lados de manzana y las unidades elementales de muestreo que en este caso serán las viviendas. Desde luego las unidades objeto de estudio son las personas que habitan las viviendas; el procedimiento de selección consiste en seleccionar en forma aleatoria al interior de cada estrato un conjunto de lados de manzana y dentro de cada lado seleccionado obtener a su vez una muestra de viviendas en forma aleatoria. Se presentan y explican los estimadores tanto del parámetro investigado como de la varianza de dicho estimador presentando también la expresión del tamaño de muestra y la forma de repartirla (afijación) en los diferentes estratos considerados en la población investigada. (Vivanco, 2005).

2. Muestreo e Investigación de Mercados

El muestreo es una metodología fundamental en cualquier estudio de mercados. Impacta a los

estudios externos, internos, los cualitativos y los cuantitativos (López, 2001). La Figura 1 hace referencia a las preguntas básicas que permiten definir correctamente los cuatro componentes básicos de un diseño muestral (Pérez, 2000).

En el mundo de hoy es muy común escuchar o leer información que hace referencia a términos como: mercado interno, mercado electoral, investigación de mercados, calidad del servicio prestado, oportunidad de inversión, diseño de productos, satisfacción al cliente, etc. Así mismo, hay gran cantidad de datos que hablan del consumo de productos; tales como los productos alimenticios, el mercado de automóviles, la canasta familiar, consumo de gaseosa, etc. Estos datos tienen como base una metodología científica basada en una estructura de muestreo que permita encontrar estimaciones estadísticas confiables de tal manera que se pueda controlar la incertidumbre presente al generalizar y obtener conclusiones sobre la población analizada, de la cual se ha extraído la muestra.

3. El Modelo de Muestreo Bietápico

En este apartado se caracteriza el modelo de muestreo bietápico estratificado (Cochran, página 386) para la estimación de parámetros asociados al estudio de variables cualitativas.

Una variable cualitativa es aquella que por su naturaleza al ser observada o medida no produce como resultado un número siendo este simplemente un nombre; así por ejemplo la condición de si trabaja o no para un encuestado puede representarse respectivamente por: 1=Si trabaja y 0=No trabaja; o por cualquier par de números que el investigador determine. Al estudiar variables cualitativas el resultado de las mismas cuando se realiza una investigación se representa por medio de una proporción o porcentaje como por ejemplo: "Porcentaje de jefes de hogar que trabajan". El modelo bietápico hace referencia a que las unidades objeto de estudio (población investigada) se encuentran agrupadas o sea formado conjuntos de diferente cantidad de ellas (Ospina, pagina 191) como es el caso de los



Figura 1. Preguntas básicas para los componentes de un diseño muestral

jefes de hogar de una ciudad como Cali, para los cuales; cada uno de ellos pertenece a una vivienda, la cual a su vez pertenece a un lado de manzana que en este caso llamaremos conglomerado. Por otro lado es sabido que el comportamiento de los consumidores depende del estrato socioeconómico al cual pertenecen y en este sentido considerar el lado de manzana es de suma utilidad ya que en las principales ciudades La Oficina de Planeación Municipal tiene caracterizados todos los lados de manzana que

componen la ciudad, estableciendo el estrato socioeconómico al cual pertenece cada uno de ellos.

El modelo propuesto recoge entonces de un lado, la forma de presentación de las unidades de estudio (grupos) y de otro la condición de comportamiento frente al consumo de un bien en el sentido de que dicho comportamiento es más parecido entre los hogares que pertenecen al mismo estrato socioeconómico y muy diferente

entre hogares de diferente estrato. La condición de bietápico hace referencia a que una vez seleccionado en forma aleatoria un lado de manzana se procede a seleccionar en forma aleatoria una muestra de viviendas de dicho lado; este procedimiento debe realizarse en la muestra de lados de manzana de cada estrato socioeconómico.

▪ **Parámetro investigado y su estimador asociado**

El parámetro investigado es la proporción de viviendas en las cuales se consume determinado producto en una determinada ciudad es decir:

$$P = \frac{A}{N}$$

En donde: A representa el total de viviendas en donde se consume el producto y N será el total de viviendas en la ciudad investigada. El estimador para dicho parámetro en el modelo referido tendrá la siguiente estructura:

$$\begin{aligned}\hat{P} &= \sum_{h=1}^L W_h \hat{P}_h = \sum_{h=1}^L W_h \left(\frac{\sum_{i=1}^{n_h} \hat{A}_{hi}}{\sum_{i=1}^{n_h} M_{hi}} \right) \\ &= \sum_{h=1}^L W_h \left(\frac{\sum_{i=1}^{n_h} M_{hi} \cdot \hat{P}_{hi}}{\sum_{i=1}^{n_h} M_{hi}} \right); \quad \hat{P}_{hi} = \frac{a_{hi}}{m_{hi}}\end{aligned}$$

En donde:

a_{hi} : Número de viviendas en la muestra de viviendas del lado i-ésimo del estrato h-ésimo de la ciudad en las cuales se consume el producto.

m_{hi} : Número de viviendas muestreadas en el lado i-ésimo del estrato h-ésimo de la ciudad.

n_h : Número de lados muestreados en el estrato h-ésimo de la ciudad.

M_{hi} : Número de viviendas que componen el lado i-ésimo del estrato h-ésimo de la ciudad.

W_h : Peso del estrato h representado por la siguiente expresión:

$$W_h = \frac{N_h \bar{M}_{0h}}{\sum_{h=1}^L N_h \bar{M}_{0h}}$$

En donde:

N_h : Número de lados de manzana del estrato h habitados.

$\bar{M}_{0h}$: Promedio de viviendas por lado de manzana del estrato h.

Nota:

Generalmente no se conoce el promedio $\bar{M}_{0h}$; en este caso se estima mediante:

$$\hat{\bar{M}}_{0h} = \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi}$$

▪ **Varianza del estimador y componentes-
Estimador de la varianza del estimador**

La varianza del estimador propuesto anteriormente viene dada por:

$$V(\hat{P}) = \sum_{h=1}^L W_h^2 V(\hat{P}_h)$$

$$V(\hat{P}_h) = \frac{1}{M_{0h}^2} \left\{ \frac{N_h(N_h - n_h)}{n_h} S_{UPM(h)}^2 + \frac{N_h}{n_h} \sum_{i=1}^{N_h} M_{hi}(M_{hi} - m_{hi}) \frac{S_{2hi}^2}{m_{hi}} \right\}$$

$$S_{UPM(h)}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (A_{hi} - M_{hi}P_h)^2$$

$$P_h = \frac{\sum_{i=1}^{N_h} A_{hi}}{\sum_{i=1}^{N_h} M_{hi}}$$

$$S_{2hi}^2 = \frac{M_{hi}P_{hi}(1 - P_{hi})}{M_{hi} - 1}$$

$$P_{hi} = \frac{A_{hi}}{M_{hi}}$$

En donde:

P_{hi} : Proporción poblacional de viviendas en las cuales se consume el producto analizado para el lado i-ésimo del estrato h-ésimo de la ciudad investigada.

A_{hi} : Número total de viviendas en las cuales se consume el producto analizado para el lado i-ésimo del estrato h-ésimo de la ciudad investigada.

P_h : Proporción poblacional de viviendas en el estrato h-ésimo de la ciudad analizada en las cuales se consume el producto analizado. (razón al tamaño) (Scheaffer et al, 2007).

S_{2hi}^2 : Variabilidad al interior del lado i-ésimo del estrato h-ésimo de la ciudad analizada respecto a si se consume o no el producto analizado.

$S_{UPM(h)}^2$: Variabilidad entre lados del estrato h-ésimo de la ciudad analizada respecto a si se consume o no el producto analizado.

La varianza estimada del estimador de la proporción de viviendas en las cuales se consume el producto analizado está definida de la siguiente manera:

$$\hat{V}(\hat{P}) = \sum_{h=1}^L W_h^2 \hat{V}(\hat{P}_h)$$

$$\hat{V}(\hat{P}_h) = \frac{1}{M_{0h}^2} \left\{ \frac{N_h(N_h - n_h)}{n_h} \hat{S}_{UPM(h)}^2 + \frac{N_h}{n_h} \sum_{i=1}^{n_h} (M_{hi} - m_{hi}) \frac{\hat{S}_{2hi}^2}{m_{hi}} \right\}$$

$$\hat{S}_{UPM(h)}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (\hat{A}_{hi} - M_{hi}\hat{P}_h)^2$$

$$\hat{A}_{hi} = \sum_{i=1}^{n_h} M_{hi} \cdot \hat{P}_{hi}$$

$$\hat{S}_{2hi}^2 = \frac{M_{hi}\hat{P}_{hi}(1 - \hat{P}_{hi})}{m_{hi} - 1}$$

4. Obtención de la Expresión para el Tamaño de Muestra Requerido al Usar el Modelo

▪ Nivel de confianza establecido

El nivel de confianza estará representado por el valor del percentil $(1-\alpha)100\%$ bajo la distribución normal. Por lo regular se asumen niveles de confianza superiores al 90%. En lo sucesivo el nivel de confianza estará representado por: $Z_{\alpha/2}$

▪ Error de muestreo

El error de muestreo está definido como la máxima discrepancia entre el valor dado por el estimador de P y su correspondiente parámetro (investigado) es decir: $EM = |\hat{P} - P|$. No se acostumbra que el error cometido sea superior al 10% del parámetro investigado.

Nota: Para asumir un error razonablemente cuando no se tiene conocimiento previo de este, se realiza un estudio piloto y se tiene en forma adelantada el valor de P y con este se procede a calcular el EM en términos porcentuales. Es decir un porcentaje inferior al 10% del valor de P estimado en el estudio piloto.

▪ Tamaño de muestra

La expresión del tamaño de muestra se obtiene de la expresión de varianza dada en el apartado anterior; Teniendo en cuenta las siguientes consideraciones:

1. Se hace que la representación del número de viviendas al interior de cada lado de manzana en cada estrato sea proporcional al número de ellas en el lado analizado (autoponderación) para ello suponemos constante la fracción de

muestreo en la segunda etapa para cada estrato; es decir

$$f_{2hi} = \frac{m_{hi}}{M_{hi}} = \frac{\bar{m}_h}{\bar{M}_h} = \text{constante} = f_{2h}$$

$$\Leftrightarrow m_{hi} = f_{2h} \cdot M_{hi}$$

Tanto $\bar{m}_h$ como $\bar{M}_h$ serán obtenidas sobre la base de el estudio piloto realizado.

2. Se supone que la forma como serán repartidos los lados de manzana entre los estratos socioeconómicos (Clairin, et al., 1997) de la ciudad será proporcional a la cantidad de viviendas en los lados de manzana del estrato correspondiente, es decir realizamos la siguiente asignación:

$$W_h = (w_h) = \frac{n_h}{n} \Leftrightarrow n_h = n \cdot W_h$$

Bajo las consideraciones anteriores el tamaño de muestra requerido para estimar la proporción poblacional de viviendas en las cuales se consume determinado producto asumiendo el error de muestreo y el nivel de confianza determinados viene dado por la expresión:

$$n \geq \frac{\sum_{h=1}^L W_h (B_h + C_h)}{V(\hat{P}) + \sum_{h=1}^L W_h^2 \left(\frac{B_h}{N_h} \right)}$$

$$B_h = \left(\frac{S_{UPM(h)}^2}{\bar{M}_{0h}^2} \right)$$

$$C_h = \frac{\sum_{i=1}^{N_h} M_{hi} (1 - f_{2h}) \cdot S_{2hi}^2}{(M_{0h}) \bar{m}_h}$$

En donde:

$V(\hat{P})$: Será la varianza deseada la cual es obtenida mediante la expresión:

$$V(\hat{P}) = \left(\frac{EM}{Z_{\alpha/2}} \right)^2$$

Nota: Las expresiones B_h y C_h deben ser

estimadas por adelantado mediante el estudio piloto.

5. Protocolo para la Obtención de las Unidades que serán Medidas o Encuestadas

La Figura 2 muestra la secuencia que se utilizaría para la obtención de las unidades que serán medidas o encuestadas finalmente.

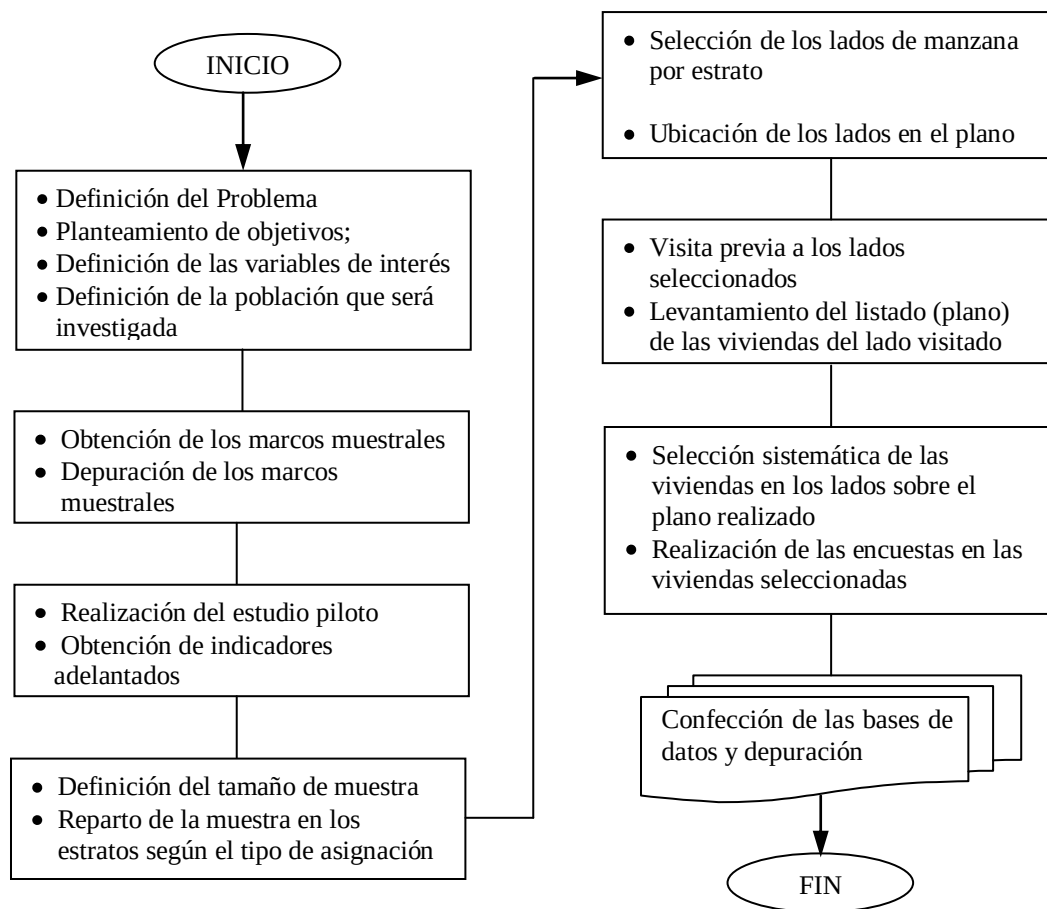


Figura 2. Protocolo de la obtención de las unidades que serán medidas.

6. Conclusiones y Recomendaciones

Las siguientes son las recomendaciones para la aplicabilidad del modelo de muestreo antes expuesto:

- Tener claro el problema que se desea investigar y plantearlo en forma de objetivos medibles.
- Definir muy claramente la(s) proporción(es) que serán estimadas y tener claro la forma de medirlas en la práctica (significado en forma de cociente).
- Si se hace censo en el lado analizado se reduce la variabilidad de las estimaciones al suprimir la variabilidad debida al interior de cada conglomerado.
- El tamaño de muestra puede ser superior al tamaño planteado.
- Si la investigación plantea la estimación de más de una proporción; debe establecerse el tamaño de muestra final como aquel valor mayor (mayor n establecido) resultante de entre ellas o por lo menos las de más importantes ya que es posible que se asuman errores y niveles de confianza diferentes en cada caso.
- La conglomeración de las unidades de estudio busca reducir los desplazamientos de los entrevistadores, haciendo un mayor número de entrevistas en un mismo espacio territorial delimitado (D'Anaconda, 2004).

7. Referencias Bibliográficas

1. Clairin, Remy y Brion, Philippe. Manual de Muestreo. Editorial. La Muralla, España, 2002.
2. Cochran, William. Técnicas de muestreo: CECSA. México, 1980.
3. D'Anaconda Ma. Ángeles. Métodos de Encuesta. Teoría y Práctica, Errores y Mejora. Síntesis. España, 2004.
4. ESTADÍSTICAS AL DÍA, N° 3 marzo de 2002. Departamento Administrativo Nacional de Estadística. Bogotá.
5. Lopez, Alfredo. Investigaciones de Mercado. CECSA. México, 2001.
6. Ospina, David. Introducción al Muestreo. Universidad Nacional de Colombia. Bogotá, 2001.
7. Pérez, César. Técnicas de Muestreo Estadístico. Alfaomega. México, 2000.
8. Pérez, César. Muestreo Estadístico. Prentice Hall. España, 2005.
9. Scheaffer, Richard; Mendenhall III. William y LYMAN OTT. R. Elementos de Muestreo: Thomson; Madrid 2007.
10. Vivanco, Manuel, Muestreo Estadístico Diseño y Aplicaciones. Editorial Universitaria, Chile, 2005.