03043
3
2ej.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADEMICA DEL CICLO PROFESIONAL Y POSGRADO

INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y EN SISTEMAS

"CALCULO DEL EFECTO DE DISEÑO PARA ENCUESTAS ELECTORALES A NIVEL MUNICIPAL"

T E S I N A

QUE PARA OBTENER EL DIPLOMA DE

ESPECIALIZACION EN

ESTADISTICA APLICADA

P R E S E N T A:

MONICA TINAJERO BRAVO

MAYO 1998

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADEMICA DEL CICLO PROFESIONAL Y POSGRADO

INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y EN SISTEMAS

"CALCULO DEL EFECTO DE DISEÑO PARA ENCUESTAS ELECTORALES A NIVEL MUNICIPAL"

T E S I N A

QUE PARA OBTENER EL DIPLOMA DE

ESPECIALIZACION EN ESTADISTICA APLICADA

PRESENTA

MONICA TINAJERO BRAVO

MAYO 98

DEDICATORIAS:

A mis papás:

Porque con su ejemplo y amor me enseñaron el camino.

A mis hermanos Ricky, Nando y Brendis:

Por su cariño y por hacerme parte del equipo.

A Abi:

Por su amistad incondicional.

A Norma y Gaby:

Por todos los momentos que compartimos.

A Paco:

Por su apoyo y comprensión, con todo mi amor.

AGRADECIMIENTOS:

A la Dra. Guillermina Eslava G. por la asesoría brindada para la realización de este trabajo. Muchas gracias.

Al Dr. Ignacio Méndez, E.E.A. Catalina Palmer, M. en C. Martín Romero y M. en C. Salvador Zamora por sus valiosos comentarios y el tiempo que le dedicaron al presente trabajo.

A Helen por su apoyo y oportunos comentarios.

A la U.N.A.M y a mis profesores. Gracias por las enseñanzas y experiencias transmitidas.

A mis amigos y compañeros de escuela, gracias por el apoyo y la convivencia.

A la empresa Consulta, S.A. de C.V., por la información proporcionada para llevar a cabo esta tesina.

CONTENIDO

INTRODUCCION

La inquietud del ser humano por conocer y comprender del entorno que le rodea le lleva a observar, estudiar, analizar diversos fenómenos y muchas veces a expresar de manera cuantitativa algunos de ellos. En nuestro país, recientemente se ha incrementado el uso y difusión de información estadística, ejemplos de ello son el número de habitantes en determinada área geográfica, el número promedio de hijos, la opinión de los ciudadanos sobre si el Presidente de la República debe elegir al regente del D.F. o debe ser elegido por voto, la distribución del voto por partido en determinada elección, etc. Para generar medidas o indicadores que resuman tal información, se puede recurrir a los censos (que conllevan elevados costos y tiempo de realización) o bien a una encuesta por muestreo que sólo recaba datos de una muestra, es decir, de una parte de la población.

El principal objetivo de la estadística es hacer inferencias con base en una muestra, y es aquí donde surge una pregunta: ¿cómo elegir esta muestra? La respuesta no es única y depende de diversos factores. La teoría del muestreo nos sugiere cómo elegir la muestra de la mejor manera; asimismo la selección habrá que confrontarla con los recursos disponibles y llegar a una conciliación entre objetivos y recursos.

Dada la importancia que las encuestas electorales han adquirido en México en épocas recientes, en el presente trabajo se pretende comparar diferentes esquemas de muestreo en un contexto muy específico: el de conteos rápidos efectuados en algunos municipios del Estado de México el pasado 10 de noviembre de 1996.

Así entonces el propósito de esta tesina es comparar posibles diseños de muestreo en términos de diversos factores como la varianza del estimador, el sesgo y el efecto de diseño. Se le da especial importancia a este último porque se ha incrementado su uso con el desarrollo de programas computacionales para muestreo, así como por su utilidad práctica

Si bien es cierto que el análisis sólo se hace para una población particular y las conclusiones no se pueden generalizar, se espera que este trabajo sirva como apoyo para la planeación de futuros diseños de muestreo para conteos rápidos.

RESUMEN

El objetivo de este trabajo es comparar los esquemas de muestreo factibles, con base en conteos rápidos efectuados en siete municipios durante las elecciones para Presidentes Municipales en el Estado de México el pasado 10 de noviembre de 1996. Para cumplir con este propósito la tesina se dividió en tres capítulos.

El primero de ellos se refiere a los principales diseños de muestreo: aleatorio simple, estratificado, por conglomerados y sistemático. En él se describe cada esquema, sus principales ventajas y limitaciones, así como sus estimadores. Dado el objetivo del trabajo, se dedica un apartado al efecto de diseño, cuya definición permite comparar la eficiencia entre diversos tipos de muestreo, con respecto a un muestreo aleatorio simple. Por último, se hace una diferenciación entre los errores de muestreo y no muestrales con base a las fuentes que los originan.

En el segundo capítulo se habla de las encuestas electorales en particular. Primero se presentan algunos antecedentes y se describen brevemente los principales tipos: encuestas preelectorales y post electorales, encuestas postvoto (del inglés *exit poll*) y conteos rápidos. Posteriormente se describen algunos aspectos sobre los comicios en los que se basa esta tesina; para ello se expone un perfil electoral de los dominios de estudio[1], luego se describe el diseño de muestreo en el que se basó una empresa encuestadora, así como los tamaños de muestra que utilizó y las estimaciones obtenidas.

En el tercer capítulo se hace un análisis comparando los diseños posibles en este caso particular. La comparación involucra diversos elementos como el sesgo de los estimadores, su varianza, el efecto de diseño y la eficiencia relativa. Estas cantidades se pudieron calcular porque para cada municipio, se contó con los resultados oficiales desagregados a nivel de sección.

Finalmente, se presentan las conclusiones que se desprenden de este trabajo y dos anexos: el primero de ellos se refiere a la deducción de la varianza cuando se utiliza un estimador de razón, y en el segundo, se presentan los estimadores que se utilizaron para obtener los resultados del tercer capítulo.

---

[1] Los dominios de estudio correspondieron a los municipios de: Atlacomulco, Ixtlahuaca, Jilotepec, Nezahualcóyotl, Tlalnepantla, Toluca y Valle de Chalco

1. PRINCIPALES DISEÑOS DE MUESTREO

La realización de una encuesta implica diversas etapas, todas ellas son de importancia y si alguna falla esto puede repercutir en toda la encuesta. Estas se pueden resumir en:

- Determinación de objetivos
- Definición de la población objeto de estudio
- Grado de precisión deseado
- Métodos de medición
- Diseño de muestreo
- Posibilidad de efectuar una encuesta piloto
- Operativo de campo
- Procesamiento, análisis y presentación de resultados

Estas etapas no son las únicas, aunque diversos autores señalan que son las más importantes. Tampoco son independientes entre sí, de hecho, debe existir una retroalimentación, y un trabajo pobre o mal realizado en cualquiera de estas puede arruinar toda la encuesta.

El diseño de muestreo es sólo una parte del diseño de encuestas: "el propósito de la teoría del muestreo es que éste sea más eficiente. Su objetivo es desarrollar métodos de selección de muestras y de estimación, que proporcionen, al menor costo posible, estimaciones con la suficiente exactitud para nuestros propósitos"[1]. Otra característica importante es que contiene planteamientos adecuados a poblaciones finitas. Además, los métodos de selección no están confinados únicamente a muestreo con igual probabilidad.

Si bien es cierto que el diseño de encuestas se enfoca en muestras aleatorias o probabilísticas, también existen muestras no probabilísticas. Las primeras se caracterizan porque cada elemento tiene una probabilidad conocida y no nula de ser elegido. Entre los muestreos no probabilísticos se encuentran el muestreo de juicio, en el que un experto selecciona unidades que él considera representativas; así como el muestreo de cuotas en donde los encuestadores eligen personas o elementos hasta obtener cierto número especificado (cuota)[2]; en general las cuotas se basan en variables demográficas.

En este capítulo se describirán algunos de los métodos del muestreo aleatorio. Para cada uno de ellos se darán observaciones generales, el método de selección y los principales estimadores.

---

[1] Cochran, William G. *Técnicas de Muestreo*. CECSA. México, 1990. p. 28.

[2] Kish, Leslie. *Muestreo de Encuestas*. 1ª edición. Trillas. México, 1792 p. 41, 649.

1.1. Muestreo aleatorio simple (m.a.s.)

Como su nombre lo indica es el método más elemental y consiste en seleccionar aleatoriamente n unidades de un total de N, donde cada una de las muestras de n elementos tiene la misma probabilidad de ser elegida. Para seleccionar las unidades se puede utilizar una tabla de números aleatorios o bien algún programa de computación que genere números aleatorios. Aunque en la práctica se utiliza poco, su importancia radica en:

- Posee propiedades matemáticas simples.
- Los demás diseños de muestreo se basan en este.
- Los cálculos para el m.a.s. pueden utilizarse, en algunos casos, ajustándolos por el efecto de diseño[3] correspondiente (ver apartado 1.5 de este trabajo), y así evitar cálculos de diseños más elaborados o complejos.

Dentro del m.a.s. existen dos maneras de elegir las unidades: con reemplazo y sin reemplazo. A continuación se presenta la notación y los estimadores utilizados, dependiendo si la elección de unidades elementales se hace con o sin reemplazo.

Notación:

$\bar{Y}$	Media poblacional de la característica de interés, y
σ^2	Varianza poblacional
S^2	Varianza poblacional pero con denominador $N\text{-}1$, en lugar de N[4].
N	Número de unidades en la población
y_i	Valor de la característica de interés y para la unidad i
$\bar{y}_{mas}$	Media muestral de la característica de interés y
s^2	Varianza muestral
n	Número de unidades seleccionadas

donde

$$\bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N}$$

[3] El efecto de diseño es la razón entre la varianza del diseño que se utilizó, y la varianza de un muestreo aleatorio simple.

[4] Algunos autores lo denominan "cuadrado medio poblacional".

$$S^2 = \frac{\sum_{i=1}^{N}(y_i - \overline{Y})^2}{N-1} = \frac{N}{N-1}\sigma^2$$

$$\bar{y}_{mas} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_{mas})^2}{n-1}.$$

1.1.1. Selección sin reemplazo (m.a.s.s.r.)

También se le llama muestreo irrestricto aleatorio. En este tipo de selección una unidad que ya ha sido elegida no se puede volver a elegir.

Un estimador insesgado de la media $\overline{Y}$ se obtiene mediante la expresión

$$\bar{y}_{massr} = \frac{\sum_{i=1}^{n} y_i}{n}$$

y un estimador insesgado del total Y es

$$y_{massr} = N\bar{y}_{massr} = N\frac{\sum_{i=1}^{n} y_i}{n}$$

las varianzas de la media y el total son respectivamente

$$Var(\bar{y}_{massr}) = (1-f)\frac{S^2}{n}$$

$$Var(y_{massr}) = N^2 Var(\bar{y}_{massr})$$

donde

$f = \dfrac{n}{N}$ se le conoce como fracción de muestreo.

1.1.3. Proporciones

En ocasiones se desea estimar la proporción o el porcentaje de la población que posee cierto atributo o característica. De hecho, es común que se publiquen resultados de encuestas o censos en forma de porcentajes.

Las proporciones son un caso particular de la media, donde la variable de interés sólo puede tomar los valores cero ó uno, aunque en la mayoría de los textos se manejan como un tema aparte tanto por su importancia como porque los cálculos y expresiones son más sencillos en este caso.

Dicho de otra manera, una proporción P es la media de una variable dicotómica y, donde

$$y_i = \begin{cases} 1 \text{ si el elemento } i \text{ presenta la característica de interés} \\ 0 \text{ si el elemento } i \text{ no presenta la característica de interés.} \end{cases}$$

Entonces

$$P = \frac{\sum_{i=1}^{N} y_i}{N}$$

y un estimador insesgado de la proporción P es como sigue

$$p = \frac{\sum_{i=1}^{n} y_i}{n}.$$

Si el muestreo se hace sin reemplazo, la varianza del estimador se calcula mediante

$$Var(p) = (1-f)\frac{S^2}{n}$$

y si el muestreo es con reemplazo se obtiene mediante la siguiente expresión

$$Var(p) = \frac{(N-1)}{N}\frac{S^2}{n}$$

con

$$S^2 = \frac{\sum_{i=1}^{N}(y_i - \overline{Y})^2}{N-1} = \frac{\sum_{i=1}^{N} y_i^2 - N\overline{Y}}{N-1}$$

$$= \frac{NP - NP^2}{N-1} = \frac{NP(1-P)}{N-1}$$

por lo tanto,

$$S^2 = \frac{N}{N-1}P(1-P).$$

Un estimador insesgado de *Var(p)* para el primer caso es

$$\hat{\text{var}}(p) = (1-f)\frac{s^2}{n}.$$

En el caso de muestreo con reemplazo

$$\hat{\text{var}}(p) = \frac{s^2}{n}$$

donde

$$s^2 = \frac{n}{n-1}p(1-p).$$

Dado que no se dan expresiones para porcentajes cuando se tienen otros esquemas de muestreo, el estimador para una proporción se puede obtener a partir del estimador para la media correspondiente, haciendo

$$y_k = \begin{cases} 1 & \text{si la unidad } k \text{ de la última etapa de muestreo presenta el atributo de interés} \\ 0 & \text{si la unidad } k \text{ de la última etapa de muestreo no presenta el atributo de interés.} \end{cases}$$

1.1.4. Cálculo del tamaño de muestra

Es difícil e importante en la planeación de cualquier encuesta decidir cuál será el tamaño de muestra. Una muestra grande puede ser muy costosa en cuanto a tiempo, dinero, recursos humanos, etc., en cambio, una muy pequeña puede ser pobre en precisión.

Existen diferentes aspectos que se deben considerar en el cálculo del tamaño de muestra, entre ellos se encuentran principalmente:

- Los objetivos de la encuesta. Es común que no se quiera evaluar sólo una característica de la población sino varias, e incluso, se requieran estimaciones para subpoblaciones
- Los parámetros a estimar: proporción, media, total, etc.
- El nivel de confianza deseado
- La precisión en las estimaciones, es decir, el error que se está dispuesto a aceptar
- El costo asociado

Aunque estos elementos entran en conflicto en ocasiones, se debe efectuar un balance de tal manera que se tome la decisión más adecuada. Para ello, se debe considerar la experiencia que se tiene sobre encuestas anteriores o estudios afines, la opinión de expertos en el tema, las consecuencias que producirían diferentes decisiones y los recursos de que se dispone. Lo anterior sin olvidar que la teoría del muestreo señala una manera de calcular el tamaño de muestra.

En particular, si se trata de un muestreo aleatorio simple, suponiendo que el costo se ve reflejado en el número de elementos a seleccionar n, y se desea estimar $\bar{Y}$ con una precisión deseada d y un nivel de confianza $(1-\alpha)$, es decir, si se quiere que

$$P\left[\left|\bar{y}-\bar{Y}\right|<d\right]=1-\alpha,$$

el tamaño de muestra sin considerar la corrección por finitud y suponiendo que $\bar{y}$ sigue una distribución normal se obtiene mediante

$$n_o=\frac{t_{\alpha/2}^2 S^2}{d^2}$$

finalmente, al considerar la fracción de muestreo, el tamaño de muestra requerido es

$$n=\frac{n_o}{1+n_o/N}$$

donde

$t_{\alpha/2}$	cuando la varianza es conocida, corresponde al porcentil de una distribución normal asociado al nivel de confianza deseado, y cuando no se conoce la varianza, corresponde al porcentil de una distribución t-*student* con n-1 grados de libertad.
d	es el error absoluto con respecto a la media que se está dispuesto a aceptar.

Cuando la fracción de muestreo es pequeña entonces $n_o \approx n$, y la corrección por finitud puede ignorarse.

En la práctica se desconoce S^2 generalmente, por lo que habrá que estimarla o hacer alguna suposición respecto a ella. Para ello se puede:

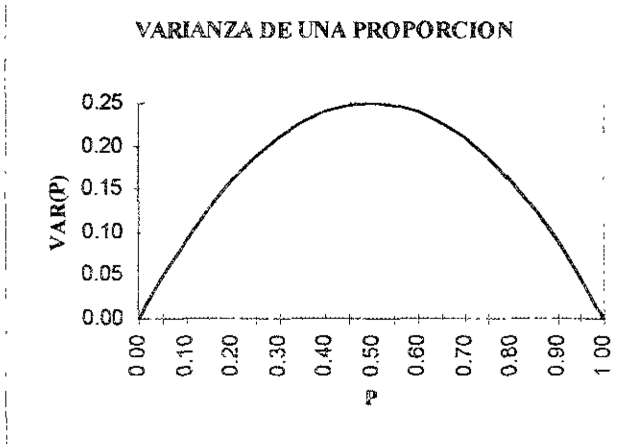
- Recurrir a encuestas anteriores o bien preguntar a expertos en el tema. En el caso que se tenga la varianza bajo un m.a.s. anterior y la muestra actual no es aleatoria simple, puede ajustarse la varianza utilizando el efecto de diseño (ver apartado 1.5.).

- Suponer el valor del coeficiente de variación C y después hacer $S = C\overline{Y}$. El coeficiente de variación de un estimador de θ es

$$C_{\hat{\theta}} = \frac{\sqrt{Var(\hat{\theta})}}{\theta}$$

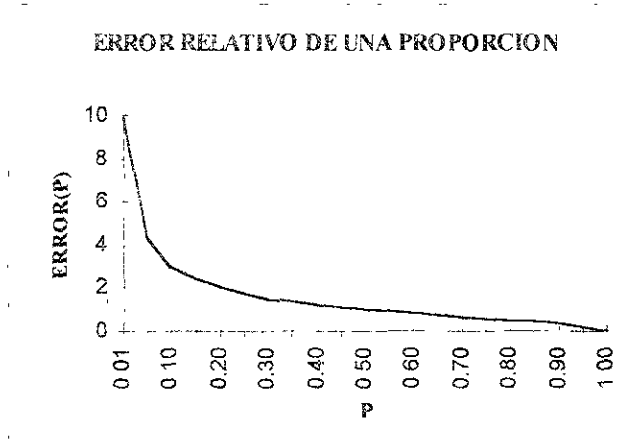
- Si se quiere estimar una proporción P, su varianza, $P(1-P)$, no es sensible a cambios en la P cuando $0.2 \leq P \leq 0.8$[5], y alcanza su máximo en $P = 0.5$ (Ver gráfica 1.1.4). Si el parámetro no pertenece a dicho rango se deberán utilizar otros métodos para calcular el tamaño de muestra, como el método de Haldane o muestreo inverso[6].

Gráfica 1.1.4.1



- Como se puede observar en la gráfica 1.1.4.2, el error relativo o coeficiente de variación de una proporción aumenta conforme esta disminuye. Así, para un error específico, el tamaño de muestra aumenta conforme la proporción es más pequeña.

Gráfica 1.1.4.2



[5] Kish, Leslie. Op. cit. p. 76.
[6] Cochran, William G. Op. cit p 109

- Efectuar una encuesta piloto.
- La varianza de la muestra no sólo depende de n sino que también depende del diseño. En diseños más elaborados o complicados que el m.a.s., denominados "diseños complejos", el tamaño de muestra se obtiene utilizando técnicas y principios análogos a los que se utilizan en el caso de un muestreo aleatorio simple, pero con los ajustes propios del esquema particular.

1.2. Muestreo aleatorio estratificado

Este diseño de muestreo consiste en dividir a los elementos de la población en grupos o subpoblaciones que no se traslapen, denominados estratos, y elegir una muestra independiente dentro de cada estrato.

Los estratos son una partición de la población, es decir, cada elemento de la población pertenece a uno y sólo un estrato. Ejemplos de estratos son: áreas socioeconómicas, estratos de marginalidad como los que maneja el CONAPO, estratos urbano y rural, etc.

Si dentro de cada estrato la muestra se selecciona mediante muestreo aleatorio simple, se dice que se tiene un muestreo aleatorio estratificado. Sin embargo, en cada estrato también se puede tener un muestreo por conglomerados, un sistemático o un diseño más complejo, e incluso, la forma de muestreo puede variar de un estrato a otro.

Se tienen más beneficios de un muestreo estratificado si los estratos son lo más homogéneos al interior y las medias entre estratos son lo más diferentes posible. En ocasiones, los estratos pueden estar dados de manera natural en la población y en otras se tienen que construir. Se tendrá mayor ganancia cuando la o las variables de estratificación estén correlacionadas con las variables objetivo. Además, se supone que se conoce el tamaño de los estratos.

Las principales razones por las que se utiliza este esquema son:

- Aumentar la precisión de las estimaciones, ya que en la mayoría de los casos el muestreo estratificado es al menos tan eficiente como el muestreo aleatorio simple
- Disminuir los costos
- Utilizar diferentes diseños de muestreo entre estratos
- Porque los grupos o subpoblaciones pueden ser además un dominio de estudio[7]

---

[7] Dominio de estudio es una parte de la población para la que se desea dar estimaciones con precisión y confianza estadísticas propias.

A continuación se dan los estimadores de la media y la varianza.

Notación:

N_h Número total de unidades en el estrato h, $\sum_{h=1}^{L} N_h = N$

W_h Peso o ponderación del estrato h, $W_h = N_h / N$

$\bar{y}_h$ Media muestral del estrato h

$Var(\bar{y}_h)$ Varianza de la media muestral del estrato h

$\hat{\mathrm{var}}(\bar{y}_h)$ Estimador de la varianza de la media muestral del estrato h

Un estimador insesgado de la media poblacional es

$$\bar{y}_{est} = \sum_{h=1}^{L} W_h \bar{y}_h$$

cuya varianza está dada por

$$Var(\bar{y}_{est}) = \sum_{h=1}^{L} W_h^2 Var(\bar{y}_h)$$

y su estimador insesgado es

$$\hat{\mathrm{var}}(\bar{y}_{est}) = \sum_{h=1}^{L} W_h^2 \ \hat{\mathrm{var}}(\bar{y}_h)$$

donde

$$W_h = \frac{N_h}{N}.$$

Si se trata de un muestreo aleatorio estratificado, los estimadores para cada estrato se obtienen mediante las fórmulas que se dan en los incisos 1.1.1 y 1.1.2., en el caso particular de que la media sea una proporción se utilizan los estimadores del apartado 1.1.3.

En cuanto a la forma de calcular el tamaño de muestra, existen métodos para llevarlo a cabo de manera óptima si se tiene un costo fijo o bien una precisión deseada[8]. Otra alternativa es calcular el tamaño de muestra para un m.a.s. y después ajustarlo por el efecto de diseño.

[8] Sukhatme, P.V et. al. *Sampling Theory of Surveys with Applications*. 3ª edición, Iowa State University Press. Iowa, 1984, p.107.

1.3. Muestreo por conglomerados en una etapa

El muestreo por conglomerados es aquel diseño en el que la unidad de selección es un conjunto o conglomerado de elementos de la población. Los conglomerados son una partición de esta, es decir, cada elemento de la población pertenece a uno y sólo un conglomerado. El número de elementos puede variar de conglomerado a conglomerado.

Ejemplos de conglomerados son los municipios (conglomerados de colonias, manzanas, viviendas), las familias (conglomerados de personas), las secciones electorales (conglomerados de electores, votantes), etc.

Este tipo de muestreo se utiliza porque:

- El costo por elemento es menor.
- Es más fácil o menos costoso obtener un marco para los conglomerados que resulten seleccionados que para todos los elementos de la población. De hecho, en algunas ocasiones, es prácticamente imposible tener un marco con las unidades elementales.
- Los conglomerados están dados de manera natural en la población.

En contra parte, posee las siguientes desventajas:

- La varianza por elemento es mayor, esto es debido a que, en general, hay homogeneidad dentro de los conglomerados.
- Existe mayor dificultad en el análisis estadístico.

A diferencia con el muestreo estratificado, es deseable que la heterogeneidad dentro de los conglomerados sea lo más grande posible y lo más pequeña entre conglomerados.

Como se mencionó anteriormente, la unidad de selección es el conglomerado. Sin embargo, existen diferentes maneras de seleccionarlo: con igual probabilidad o con diferente, con reemplazo o sin reemplazo. Además, dentro de cada conglomerado se puede hacer un censo, que generalmente se denomina "selección de conglomerados en una etapa", o bien, un submuestreo de unidades del conglomerado denominado "selección de unidades en dos o más etapas". En este trabajo sólo se describirá el de una etapa.

1.3.1. Muestreo de conglomerados con igual probababilidad

Se supone que los conglomerados son seleccionados con un muestreo aleatorio simple y en la muestra se incluyen todos los elementos de los conglomerados seleccionados. El muestreo de unidades primarias puede ser con o sin reemplazo. Análogamente al m.a.s., la diferencia entre los estimadores de ambos casos radica principalmente en el factor de corrección por finitud, como se verá cuando se presenten las fórmulas para la varianza.

En la práctica, generalmente los conglomerados tienen tamaños diferentes, lo cual implica que el tamaño de la muestra sea una variable aleatoria. Por ende, unos de los estimadores de la media y el total involucran un cociente de variables aleatorias. Existen otros estimadores para el total y para la media[9], cuya precisión es pobre cuando los totales por conglomerado varían poco entre unidades y los tamaños de los conglomerados varían considerablemente[10]. Al estimador del cociente de dos variables también se le llama de razón y su expresión se muestra a continuación (para mayor detalle ver el Anexo A).

Notación:

M Total de conglomerados en la población

m Número de conglomerados en la muestra

N_i Tamaño del conglomerado i, $\sum_{i=1}^{M} N_i = N$

Y_i Total de la característica de interés y, en el conglomerado i, $\sum_{i=1}^{M} Y_i = Y$

R Media poblacional, $R = \frac{Y}{N}$

Cuando se conoce el tamaño de cada conglomerado, N_i, un estimador[11] de R es el siguiente

$$r = \frac{y'}{N'_m} = \frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{m} N_i} = \frac{M/m\sum_{i=1}^{m} Y_i}{M/m\sum_{i=1}^{m} N_i} = \frac{y}{N_m}$$

donde:

$y = M/m\sum_{i=1}^{m} Y_i$ es un estimador insesgado de Y

$N_m = M/m\sum_{i=1}^{m} N_i$ es un estimador insesgado de N

y una aproximación a su varianza[12], derivada empleando series de Taylor (ver anexo A) es la siguiente

$$Var(r) \doteq \frac{1}{N^2}(Var(y) + R^2 Var(N_m) - 2RCov(y, N_m)).$$

[9] Ibid. p. 291.

[10] Cochran, William G. Op. cit. p. 310.

[11] Hansen, Morris H. et. al. *Sample Survey Methods and Theory Vol. I.* John Willey & Sons. New York, 1953 p. 253.

[12] Kish, Leslie. Op. cit. p. 248.

Cuando se eligen conglomerados sin reemplazo las varianzas y covarianza[13] de los estimadores del numerador y denominador están dados por

$$Var(y) = \frac{M^2(M-m)}{M} \frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{m(M-1)}$$

$$Var(N_m) = \frac{M^2(M-m)}{M} \frac{\sum_{i=1}^{M}(N_i - \frac{N}{M})^2}{m(M-1)}$$

$$Cov(y, N_m) = \frac{M^2(M-m)}{M} \frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{m(M-1)}$$

sustituyendo y simplificando se llega a

$$Var(r) \doteq \frac{M^2(M-m)\sum_{i=1}^{M}(Y_i - N_i R)^2}{N^2 Mm(M-1)},$$

expresión presentada también por Cochran[14].

Un estimador de la varianza[15] es

$$\hat{var}(r) \doteq \frac{m}{m-1} \frac{(M-m)\sum_{i=1}^{m}(Y_i - rN_i)^2}{MN_m^2}.$$

En el caso de que el muestreo de conglomerados sea con reemplazo

$$Var(y) = \frac{M\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{m}$$

[13] Hansen, Morris H. et. al. Op cit. p. 253
[14] Cochran, William G. Op cit p 310
[15] Hansen, Morris H et al Op. cit. p 257

$$Var(N_m) = \frac{M\sum_{i=1}^{M}(N_i - \frac{N}{M})^2}{m}$$

$$Cov(y, N_m) = \frac{M\sum_{i=1}^{M}(Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{m}.$$

La expresión para la aproximación de la varianza de r se reduce a

$$Var(r) \doteq \frac{M\sum_{i=1}^{M}(Y_i - N_i R)^2}{N^2 m}.$$

A pesar de que el estimador de razón es sesgado tiene la ventaja de que es consistente, es decir, su sesgo[16] se hace pequeño conforme el número de unidades primarias de muestreo m crece, y puede ser ignorado para muestras moderadamente grandes.

Una expresión aproximada para el sesgo del estimador de razón[17] es

$$Sesgo(r) = E(r) - R \doteq \frac{R\left[Var(N_m) - Cov(y, N_m)\right]}{N^2}.$$

Existe una cota superior para el sesgo[18]

$$|Sesgo(r)| \leq \sqrt{Var(r)}\,\frac{\sqrt{Var(y)}}{Y}$$

1.3.2. Muestreo de conglomerados con probabilidad proporcional al tamaño

Como su nombre lo indica, el muestreo con probabilidad proporcional al tamaño (ppt) consiste en seleccionar unidades con diferente probabilidad. Así, la probabilidad de selección de cualquier unidad es directamente proporcional a su tamaño: unidades grandes tienen mayor probabilidad de ser elegidas que unidades pequeñas.

---

[16] El sesgo de un estimador es igual al valor esperado del estimador menos el parámetro a estimar.

[17] Kish, Leslie. Op cit. p.249.

[18] Ibid. p.249.

El muestreo por conglomerados es un caso típico en el que se puede utilizar el muestreo ppt, ya que el número de elementos en el conglomerado es una medida natural del tamaño del mismo.

Se recomienda utilizarlo cuando:

- Las unidades varían en tamaño y la variable bajo estudio está correlacionada con el tamaño del conglomerado.
- Se desea controlar el tamaño de muestra. En caso de que haya submuestreo en los conglomerados se puede lograr una fracción de muestreo uniforme.

Otra ventaja es que la probabilidad de selección puede estar basada en alguna medida auxiliar o relacionada con el verdadero tamaño del conglomerado. La única condición para las probabilidades de selección es que la suma de ellas sea igual a 1.

Por otro lado, hay que tomar en cuenta que si la variación en el tamaño del conglomerado es moderada, el efecto en la varianza de la estimación de promedios o razones es pequeño. Algunos autores señalan que la reducción en la varianza como resultado de disminuir la variación en el tamaño es moderada si el coeficiente de variación del tamaño del conglomerado es menor a 0.5 ó 1.0 [19].

Los conglomerados también pueden ser seleccionados con reemplazo o sin reemplazo. Sin embargo, el primer esquema tiene la ventaja de que la probabilidad de selección de un conglomerado permanece constante de extracción a extracción a diferencia del segundo caso, en el que en cada extracción se debe de calcular la probabilidad de selección. Por esta razón sólo se verá el primero en este trabajo.

La notación es la misma que para el muestreo con igual probabilidad, sólo hay que agregar:

π_i Probabilidad de selección del conglomerado i

Un estimador de la media[20] R esta dado por

$$r_{ppt} = \frac{y'_{ppt}}{N'_{m,ppt}} = \frac{\sum_{i=1}^{m} Y_i / \pi_i}{\sum_{i=1}^{m} N_i / \pi_i} = \frac{1/m \sum_{i=1}^{m} Y_i / \pi_i}{1/m \sum_{i=1}^{m} N_i / \pi_i} = \frac{y_{ppt}}{N_{m,ppt}}$$

[19] Hansen, Morris H. et. al. Op. cit. p 354.
[20] Ibid. p. 360.

donde:

$$y_{ppt} = 1/m\sum_{i=1}^{m} Y_i \Big/ \pi_i \qquad \text{es un estimador insesgado de } Y$$

$$N_{m,ppt} = 1/m\sum_{i=1}^{m} N_i \Big/ \pi_i \qquad \text{es un estimador insesgado de } N.$$

Una aproximación a su varianza es la siguiente

$$Var(r_{ppt}) \doteq \frac{1}{N^2}[Var(y_{ppt}) + R^2 Var(N_{m,ppt}) - 2RCov(y_{ppt}, N_{m,ppt})]$$

donde las varianzas y covarianzas[21] indicadas en la ecuación anterior, se calculan por

$$Var(y_{ppt}) = \frac{\sum_{i=1}^{M} \pi_i (\frac{Y_i}{\pi_i} - Y)^2}{m}$$

$$Var(N_{m,ppt}) = \frac{\sum_{i=1}^{M} \pi_i (\frac{N_i}{\pi_i} - N)^2}{m}$$

$$Cov(y_{ppt}, N_{m,ppt}) = \frac{\sum_{i=1}^{M} \pi_i (\frac{Y_i}{\pi_i} - Y)(\frac{N_i}{\pi_i} - N)}{m}.$$

Al igual que en el muestreo con igual probabilidad, el estimador de razón es sesgado, cuyo sesgo está dado por una expresión análoga al de igual probabilidad, pero $Var(N_m)$ y $Cov(y, N_m)$ se calculan según las fórmulas anteriores.

Un estimador de la varianza es

$$\hat{var}(r_{ppt}) \doteq \frac{1}{N_{m,ppt}^2}(\hat{var}(y_{ppt}) + r_{ppt}^2\, \hat{var}(N_{m,ppt}) - 2r_{ppt}\, \hat{cov}(y_{ppt}, N_{m,ppt}))$$

donde

$$\hat{var}(y_{ppt}) = \frac{\sum_{i=1}^{m} (\frac{Y_i}{\pi_i} - y_{ppt})^2}{m(m-1)}$$

[21] Ibidem. p. 360.

$$\hat{\text{var}}(N_{m,ppt}) = \frac{\sum_{i=1}^{m}(\frac{N_i}{\pi_i} - N_{m,ppt})^2}{m(m-1)}$$

$$\hat{\text{cov}}(y_{ppt}) = \frac{\sum_{i=1}^{m}(\frac{Y_i}{\pi_i} - y_{ppt})(\frac{N_i}{\pi_i} - N_{m\,ppt})}{m(m-1)}.$$

1.3.3. Muestreo estratificado de conglomerados

Como se vio anteriormente, la estratificación puede reducir la varianza. Por ello, también se pueden hacer un muestreo estratificado de conglomerados, que combina muchas de las ventajas y limitaciones del muestreo estratificado y de conglomerados.

Cuando se tiene este diseño, existen dos estimadores de razón: el estimador de razón combinado y el estimador separado.

El estimador de razón combinado, se obtiene mediante la suma de la variable respectiva sobre todos los estratos. A continuación se muestra el estimador de la media[22]

$$r = \frac{y}{N_m} = \frac{\sum_{h=1}^{H}\frac{M_h}{m_h}\sum_{i=1}^{m_h} Y_{hi}}{\sum_{h=1}^{H}\frac{M_h}{m_h}\sum_{i=1}^{m_h} N_{hi}} = \frac{\sum_{h=1}^{H} y_h}{\sum_{h=1}^{H} N_{m_h}}$$

donde:

- h representa el estrato
- y_h es un estimador insesgado del total Y_h en el estrato
- N_{m_h} es un estimador insesgado del total N_h en el estrato

Si la selección de los conglomerados fue aleatoria o mediante ppt, los estimadores están dados por las fórmulas señaladas en los apartados 1.3.1 y 1.3.2, respectivamente.

La varianza del estimador[23] de razón se obtiene por

$$Var[r] \doteq \frac{1}{N^2}(Var[y] + R^2 Var[N_m] - 2RCov[y, N_m])$$

[22] Ibidem. p. 316, 417

[23] Kish, Leslie. Op. cit. p.231.

ya que la selección es independiente entre estratos:

$$Var[y] = \sum_{h=1}^{H} Var(y_h)$$

$$Var[N_m] = \sum_{h=1}^{H} Var(N_{m_h})$$

$$Cov[y, N_m] = \sum_{h=1}^{H} Cov[y_h, N_{m_h}]$$

donde las varianzas y covarianza por estrato $Var(y_h)$, $Var(N_{m_h})$, $Cov[y_h, N_{m_h}]$ se obtienen mediante las fórmulas señaladas en los incisos 1.3.1 y 1.3.2, para lo cual es necesario considerar si la selección es con igual probabilidad o con ppt (para mayor detalle consultar el Anexo B).

En el estimador separado de razón, se calculan las razones por estrato y posteriormente se suman, ponderando por el peso del estrato correspondiente. La media estimada y la varianza de la misma[24] se obtienen mediante las siguientes expresiones

$$r = \sum_{h=1}^{H} W_h r_h = \sum_{h=1}^{H} W_h \frac{y_h}{N_{m_h}}$$

$$Var[r] = \sum_{h=1}^{H} W_h^2 Var[r_h]$$

donde la varianza de la razón por estrato, $Var[r_h]$, se calcula utilizando las fórmulas de los apartados 1.3.1. y 1.3.2.

El estimador combinado es el más utilizado porque no es necesario conocer el peso de los estratos, es computacionalmente más sencillo, y su sesgo es menor al estimador separado. Sin embargo, este último puede tener una varianza menor cuando las medias por estrato varían considerablemente, aunque presenta el incoveniente que el sesgo de las medias por estrato puede ser grande en comparación con sus varianzas.

[24] Ibid cit. p.246.

1.4. Muestreo sistemático

Este tipo de muestreo es un procedimiento de selección, el cual consiste en dividir el tamaño total de la población N en n zonas con k unidades, donde $k=N/n$, posteriormente se elige un número aleatorio a entre 1 y k. Así, la primera unidad seleccionada es la a, la segunda es $a+k$, la tercera $a+2k$ y así sucesivamente hasta tener n unidades en muestra.

El muestreo sistemático corresponde al muestreo por conglomerados, donde el tamaño de los conglomerados es n y el número de conglomerados seleccionados es uno.

Las ventajas de este método son:

- Facilidad para obtener la muestra.
- En ocasiones evita errores, sobre todo en campo.
- Parece ser más preciso que el m.a.s. ya que refleja cualquier estratificación que exista en el ordenamiento del marco de muestreo.

A pesar de su conveniencia práctica, teóricamente las varianzas presentan un problema: "en un sentido estricto, una selección sistemática de unidades no es medible, porque la varianza no puede calcularse a partir de la muestra solamente"[25], esto se debe a que el muestreo sistemático corresponde a la selección de un conglomerado, como se mencionó anteriormente, y para calcular la varianza se necesitan al menos dos conglomerados. Sin embargo, si el orden de la lista es aleatorio con respecto a la variable que se está midiendo, entonces una muestra con selección sistemática puede aceptarse como una buena aproximación a una selección aleatoria[26]. Por lo tanto, para el cálculo de varianzas se pueden utilizar las fórmulas vistas en este capítulo.

Si existen tendencias monótonas o fluctuaciones periódicas en la lista de la población, la efectividad del muestreo sistemático depende del tamaño de la zona, k, y la longitud del periodo. Así, en el caso menos favorable, cuando se eligen observaciones iguales, la varianza estimada puede dejar de incluir fuentes de variación; y en el caso más favorable, cuando la media de cada muestra sistemática es igual a la media poblacional, se puede tener una varianza muestral nula[27].

Además, existen otros métodos para calcular aproximaciones a la varianza[28], entre los que se encuentran:

- Considerar el muestreo sistemático como un muestreo aleatorio estratificado seleccionando dos unidades por cada estrato de $2k$ unidades. Bajo este método se han propuesto diferentes estimadores de la varianza, los cuales consideran diferencias sucesivas entre las unidades en un mismo estrato.

---

[25] Ibidem. p. 148.
[26] En ocasiones a este tipo de muestreo se le llama *pseudoaleatorio*. Kish, Leslie. Op. cit. p.152.
[27] Cochran, William G. Op. cit. p. 271.
[28] Sukhatme, P.V. Op. cit p. 427.
Wolter, K. M. *Introduction to Variance Estimation*. Springer-Verlang. New York, 1985. p. 248

- Dividir a la muestra seleccionada en submuestras de igual número de unidades, con lo que se tendría más de un conglomerado.

- Otra clase de estimadores se deriva de varios supuestos acerca de la correlación entre unidades sucesivas en la población.

No obstante, para elegir el estimador más apropiado se deben considerar su sesgo, su error cuadrático medio, propiedades, etc., así como también conocer la población objeto de estudio.

1.5. Efecto de Diseño

El efecto de diseño es la razón entre la varianza de un estimador bajo un esquema de muestreo dado y la varianza de un muestreo aleatorio simple del mismo número de elementos, es decir, es la eficiencia relativa de un m.a.s. respecto a un muestreo dado. En general, Kish[29] ha propuesto dos definiciones del efecto de diseño[30]:

$$Deff(\hat{\theta}) = \frac{Var(\hat{\theta})}{Var(\hat{\theta}_{massr})}$$

$$Deft^2(\hat{\theta}) = \frac{Var(\hat{\theta})}{Var(\hat{\theta}_{mascr})}$$

donde

> $Var(\hat{\theta})$ es la varianza del estimador de θ bajo el esquema de muestreo que se utilizó.

Como puede observarse la diferencia entre ambas definiciones es que *Deff* es el efecto de diseño comparado con un muestreo aleatorio simple sin reemplazo, mientras que $Deft^2$ es el efecto del diseño con respecto a un muestreo aleatorio simple con reemplazo.

La segunda definición se utiliza porque el factor *(1-f)* puede ser considerado como parte del efecto de diseño y en ocasiones es difícil de calcular la fracción de muestreo *f*.

---

[29] Kish, Leslie. *Defs: why, when and how? a review*. Ann arbor: U. of Michigan, 1989.

[30] *Deff*: del inglés "*design effect*"

Cuando $Var(\hat{\theta})$, $Var(\hat{\theta}_{massr})$ o $Var(\hat{\theta}_{mascr})$ no se conocen, el efecto de diseño puede estimarse a partir de los valores muestrales de ambas cantidades, aunque para esto la varianza muestral del aleatorio simple debe expresarse en términos de la varianza estimada bajo el diseño que se empleó, lo cual resulta en ocasiones difícil de obtener.

Si el efecto es menor a uno, indica que la varianza bajo el diseño utilizado es menor que si se hubiera hecho un muestreo aleatorio simple. Esto casi siempre sucede, por ejemplo, cuando el diseño es estratificado. Si el efecto es mayor a uno significa que la varianza es mayor que la de un m.a.s., lo cual generalmente sucede en un muestreo por conglomerados.

Puede suceder que el estimador de θ sea sesgado pero consistente, por lo que en la práctica, cuando el sesgo es negligible, los efectos de diseño se calculan según las definiciones señaladas anteriormente[31].

El efecto de diseño se utiliza principalmente para lo siguiente:

- Calcular el tamaño de muestra. Si por ejemplo, de otra encuesta o fuente se tiene que el efecto del diseño (bajo el mismo esquema del muestreo que se quiere utilizar) es *Deff* y el tamaño de muestra necesario es n (si se tratara de un m.a.s.), entonces, el tamaño de muestra requerido, dado que se va a utilizar el esquema de muestreo complejo es:

$$n = n_{mas} Deff$$

- Economizar cálculos y presentaciones. En algunas encuestas se desean varias estadísticas para cada una de las múltiples variables que se captaron, además muchas veces esto no sólo se desea para la población total, sino también para subpoblaciones, por lo que en lugar de cada varianza o error estándar se podría presentar un efecto de diseño "general". Debe tenerse precaución en que los efectos no varíen mucho entre subpoblaciones y/o entre variables.

- Comparar la eficiencia de esquemas de muestreo complejos.

- Encuestas periódicas.

- Diseño de muestras futuras.

- "Corregir" análisis estadísticos que suponen muestras aleatorias con reemplazo.

[31] Kish, Leslie. Op. cit. p. 210.

A pesar de que el efecto de diseño es una medida fácil de interpretar y de las ventajas que posee, debe usarse con precaución y tomarse en cuenta que:

- El efecto del diseño debe ser visto como una medida "burda" para efectos grandes.
- Si las estadísticas de la encuesta son pocas, entonces es mejor calcular y presentar las varianzas para cada una de ellas.
- Muchas veces los efectos de diseño varían considerablemente de variable a variable, e incluso, pueden ser muy diferentes entre las subpoblaciones de una misma variable.
- En muestreo por conglomerados el efecto del diseño combina dos factores: la correlación intraclase y el tamaño medio del conglomerado, por lo que si se desea utilizar para encuestas futuras, se deberá tomar en cuenta el tamaño promedio de los nuevos conglomerados.

1.6. Errores de muestreo y no muestrales

Hasta el momento se ha supuesto que el error en la estimación de cierto parámetro proviene de que sólo se seleccionó una parte de la población, al cual se le denomina error de muestreo. Sin embargo, existen otros tipos de errores llamados no muestrales. Como su nombre lo indica, son errores cuyas causas no están asociadas con el proceso de muestreo y pueden ocurrir aunque se lleve a cabo un censo. Este tipo de errores se originan principalmente por:

- Imperfecciones en el marco de muestreo, debido a que incluya elementos que no pertenecen a la población objeto de estudio o bien excluya algunos de ellos.
- No respuesta, es decir, cuando no se pueden obtener para cada unidad seleccionada las medidas, observaciones o respuestas requeridas en la estimación.
- Errores de medición por imperfección de los instrumentos, incapacidad del personal encargado de ello, mal cuestionario, etc.
- Errores de codificación, captura, edición, etc.

Si bien los errores de muestreo pueden ser controlados mediante el esquema de selección, los estimadores y el tamaño de muestra, los medios para controlar los errores no muestrales son muy diferentes. En la literatura[32] se proporcionan diferentes maneras de hacerlo y cada una de ellas requiere especial atención, así como cuidado y conocimiento del tema de estudio, de la población, etc.

En esta tesina sólo se consideran errores de muestreo, aunque en la planeación de cualquier encuesta se debe de considerar ambos tipos de errores y evaluar los posibles sesgos que cada tipo de error puede introducir.

[32] Foreman, E. K. *Survey Sampling Principles*. MARCEL DEKKER New York, 1991. p. 286.
Cochran, William G Op. cit. p. 435

2. ENCUESTAS ELECTORALES

2.1. Antecedentes

Como se mencionó anteriormente, en México, cada vez es más frecuente y variado el uso de encuestas. Un ámbito en el que se han aplicado es el político, en particular para conocer las preferencias electorales, opinión sobre los partidos, candidatos, etc. De hecho, las encuestas electorales se han practicado en diversos países como Gran Bretaña, E.E.U.U., Francia, y recientemente, México.

Así como el número de encuestas ha aumentado, también se ha mejorado y diversificado la metodología que se emplea en ellas. Por ejemplo, hasta 1928 los resultados de las encuestas en los Estados Unidos se daban a conocer sólo a través de los periódicos, por lo que su difusión era lenta e incompleta. Posteriormente, la radio aceleró la cobertura informativa de los comicios y, en 1952, la televisión comenzó a informar de las elecciones nacionales. La importancia llegó a tal grado, que las tres principales cadenas televisivas en ese país informaron de los resultados de las encuestas electorales. También se formó la organización llamada Servicio de Noticias de las Elecciones (*News Election Services* NES) y en 1990 se funda el *Voters Research and Survey* (VRS).

En México, las primeras encuestas a la salida de las casillas para conocer resultados electorales se efectuaron en 1990, en seis municipios conurbados del Estado de México durante las elecciones para presidentes municipales[1]. A partir de entonces se han realizado encuestas a diferentes niveles de gobierno: municipales, distritales, estatales y nacionales, pero fue hasta 1994 con las elecciones nacionales para Presidente de la República que se tuvo un mayor impulso.

2.2. Tipos de encuestas electorales

En la actualidad varias casas encuestadoras incursionan en este ramo y, dependiendo de los objetivos del estudio, practican diferentes tipos de encuestas:

- Encuestas preelectorales
- Encuestas postvoto (EP)
- Conteos Rápidos (CR)
- Encuestas postelectorales

[1] Leyva, Norma. *El Muestreo para Proporciones. Un Caso Práctico en Encuestas para medición de Preferencias Electorales Tesis.* UNAM. México, 1995 p 43

2.2.1. Encuesta preelectoral

Se realiza con anticipación a la fecha de elecciones y su objetivo es conocer la percepción de los ciudadanos hacia los candidatos, partidos, temas de campaña, etc. Generalmente, el instrumento de medición es un cuestionario y se han efectuado en las viviendas, o en la vía pública.

2.2.2. Encuesta Postvoto

La encuesta postvoto o encuesta a la salida de las casillas (del inglés *exit poll*), tiene lugar el día de la elección y consiste en entrevistar, mediante una selección "aleatoria", a los votantes afuera de las secciones electorales incluidas en la muestra, una vez que hayan emitido su voto.

Los principales objetivos de la encuesta postvoto son:

- Conocer la preferencia partidista de los votantes.
- Describir su opinión sobre diversas situaciones, así como algunas de sus características socioeconómicas.

Es importante mencionar que algunas empresas encuestadoras utilizan la estrategia de que los encuestados depositen su cuestionario en una caja cerrada, esto con la finalidad de que no se viole el principio de confidencialidad del voto de los ciudadanos.

Al igual que en una encuesta preelectoral, en la postvoto se puede incurrir en cualquiera de los siguientes errores no de muestreo:

- Registro y/o captura errónea de la información.
- Influencia del encuestador
- Limitaciones propias del cuestionario
- No respuesta
- Respuesta falsa del entrevistado

2.2.3. Conteo Rápido

Se efectúa el día de la elección, una vez que la casilla electoral cerró y los funcionarios de casilla publicaron oficialmente sus resultados en el exterior de la misma. El conteo rápido consiste en recabar el número de votos por partido publicados en las afueras de las casillas incluidas en la muestra.

Su objetivo es estimar el porcentaje de votos que obtuvieron los partidos contendientes en la campaña electoral. Dado el esquema del conteo rápido, existen menos errores no muestrales que en los dos tipos de levantamiento señalados anteriormente:

- Conteo erróneo de los votos por parte de los funcionarios de casilla
- Copia y/o captura errónea de la información.

Aunque existen menos errores no muestrales, presenta los inconvenientes de que no se tiene control sobre la hora en que se publiquen los resultados de la sección seleccionada, puede haber impugnaciones y/o anulación de casillas con fecha posterior a la del levantamiento, y, dada la naturaleza de la encuesta, no se puede dar alguna descripción de tipo socioeconómico o de opinión de los votantes. Sin embargo, su costo es menor.

2.2.4. Encuesta postelectoral

Se lleva a cabo en fecha próxima pero posterior a la de las elecciones, su objetivo es conocer percepciones, opiniones, características sociodemográficas, preferencia electoral, etc. de los ciudadanos y/o votantes. Frecuentemente, también es de interés comparar los resultados de una encuesta postelectoral con respecto a una preelectoral. Al igual que esta última, se han efectuado en viviendas o en la vía pública y se pueden tener los mismos errores diferentes a los de muestreo.

2. 3. Conteos Rápidos en algunos municipios del Estado de México

El presente trabajo se basa en las elecciones para Presidentes Municipales del Estado de México en los municipios de Atlacomulco, Ixtlahuaca, Jilotepec, Nezahualcóyotl, Tlalnepantla, Toluca y Valle de Chalco, llevados a cabo el 10 de noviembre de 1996.

Así, en este capítulo primeramente se da un breve perfil electoral de los municipios, posteriormente se describe la metodología que utilizó la empresa encuestadora, puesto que su diseño de muestreo fue retomado para la elaboración de este trabajo, y consecuentemente constituyó la fuente de información; y por último, se presentan las estimaciones para la distribución del voto.

2. 3. 1. Perfil electoral

En la planeación de toda encuesta es muy importante conocer las características de la población objetivo más relevantes a la investigación. En este trabajo se presentan algunos datos de tipo electoral sin mostrar el perfil sociodemográfico. No se incluyen las características socioeconómicas porque no se dispuso de información a nivel de sección u otra unidad de selección que pudiera ayudar en el diseño de muestreo. Sería de mayor utilidad considerar indicadores sociales y demográficos, para dominios de estudio estatales o bien nacional, si consideramos que la heterogeidad es mayor dentro de un estado o en el país que en un municipio.

Como se observa en el cuadro 2.3.1. los municipios son muy diferentes en cuanto a número de ciudadanos registrados en la lista nominal y número total de secciones. Los municipios más grandes son Nezahualcóyotl, Tlalnepantla y Toluca. También se observa que el número promedio de ciudadanos en la lista nominal varía de municipio a municipio, factor que debe tomarse en cuenta para el cálculo del tamaño de muestra.

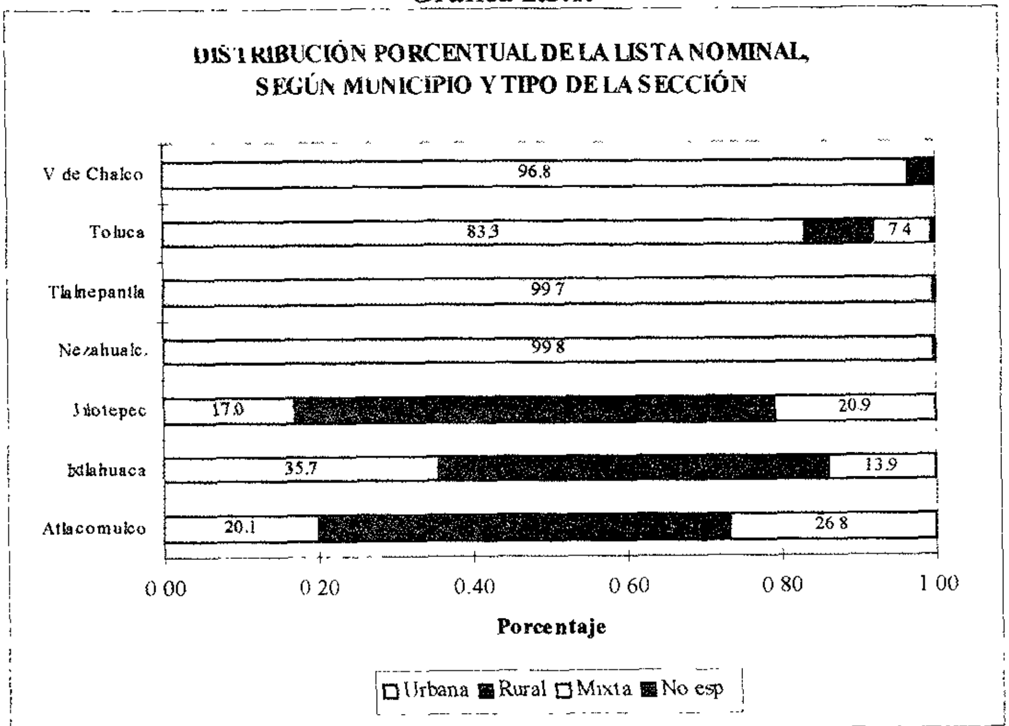
Cuadro 2.3.1
Lista nominal y número de secciones, según municipio

MUNICIPIO	*LISTA NOMINAL*[2]	*NUMERO DE SECCIONES*	*NUM. PROMEDIO DE CC. REGISTRADOS POR SECCION*
Atlacomulco	27,036	39	693
Ixtlahuaca	46,493	48	969
Jilotepec	28,520	44	648
Nezahualcóyotl	742,592	660	1,125
Tlalnepantla	411,007	362	1,135
Toluca	262,726	241	1,090
Valle de Chalco	107,306	103	1,042

Se dispuso también de información acerca del tipo de secciones que asigna el Instituto Federal Electoral. El IFE clasifica a las secciones en: urbanas, rurales, mixtas y no especificadas. En la gráfica 2.3.1 se muestra el porcentaje de la lista nominal que pertenece a cada tipo de sección. Los municipios del Area Metropolitana de la Ciudad de México están formados principalmente por secciones urbanas, mientras que en los municipios más alejados al Distrito Federal como Atlacomulco, Ixtlahuaca y Jilotepec predomina la población en secciones de tipo rural. El porcentaje de la lista nominal cuya sección no tenía un tipo especificado es pequeño, siendo el más grande de 3%.

[2]Corresponde a una lista nominal preliminar, ya que se obtuvo con anterioridad a la fecha de elección para la planeación del conteo rápido.

Gráfica 2.3.1.



2.3.2. Metodología del conteo rápido

a) Objetivo

La finalidad del conteo rápido fue estimar la distribución del voto del electorado que participaron en los comicios de Presidentes Municipales en cada uno de los municipios de: Atlacomulco, Ixtlahuaca, Jilotepec, Nezahualcóyotl, Tlalnepantla, Toluca y Valle de Chalco.

En este caso, la población objeto de estudio estuvo constituida por las personas mayores de 18 años, y que, viviendo en los municipios señalados, acudieron a las urnas el 10 de noviembre de 1996 a ejercer su derecho al voto.

Dado el objetivo, cada municipio fue un dominio de estudio cuyo marco de muestreo fue una lista de todas las secciones electorales en el municipio en estudio.

b) Diseño de muestreo

Para cada uno de los municipios o dominios de estudio, el diseño de muestreo utilizado fue estratificado, de selección de conglomerados en una etapa con probabilidad proporcional al tamaño del conglomerado, selección sistemática y con reemplazo.

La estratificación de secciones se hizo con base en la preferencia correspondiente a los comicios de agosto de 1994 para Presidente de la República. Aunque los diferentes municipios no tenían los mismos estratos necesariamente, en total se formaron cinco estratos cuya descripción se presenta en el cuadro siguiente:

Cuadro 2.3.2.1
Descripción de estratos

ESTRATO	*DESCRIPCION*[3]
1	Mayoría absoluta del PRI
2	Mayoría relativa del PRI
3	Competidos PAN-PRI o bien PAN mayoría
4	Mayoría relativa del PRI con presencia muy importante del PAN
5	Pluripartidismo PAN-PRI-PRD

Ahora bien, los estratos en los que se dividió cada municipio se muestran en el cuadro 2.3.2.2.

Cuadro 2.3.2.2
Estratos según municipio

MUNICIPIO	*ESTRATOS*
Atlacomulco[4]	1, 2
Ixtlahuaca[5]	1, 2
Jilotepec	1, 2, 5
Nezahualcóyotl	2, 4, 5
Tlalnepantla	2, 3, 4, 5
Toluca	2, 4
Valle de Chalco	1, 2, 5

[3] Los criterios numéricos para formar los estratos consideran los porcentajes para el PAN, PRI y PRD, dichos criterios se pueden resumir en.
Estrato 1 El PRI obtuvo 60% o más, el porcentaje para las otras dos fuerzas es menor a 15% La razón PRI-PAN (PRD) es 4 a 1 aproximadamente
Estrato 2 El PRI obtuvo entre 50 y 60% aproximadamente, PAN Y PRD entre 15 y 20%. La razón PRI-PAN(PRD) es 2 a 1 aproximadamente
Estrato 3 La distancia entre las dos primeras fuerzas es menor a 10 puntos y la tercera fuerza obtuvo menos de 15%, o bien, el PAN obtuvo más del 50%
Estrato 4. El PRI obtuvo entre 40 y 50% y el PAN entre 25 y 30% aproximadamente
Estrato 5. La distancia entre la primera fuerza y cada una de las otras dos es menor o igual a 15%
[4] En el diseño original Atlacomulco tenía otro estrato más, el 4, pero como era muy pequeño y el tamaño de muestra para dicho estrato sería de una sección, se unió al estrato 2.
[5] En el diseño original Ixtlahuaca tenía tres estratos, pero por el tamaño del estrato 5 se decidió colapsarlo con el estrato 2 Es importante aclarar que al igual que para Atlalcomulco, se buscó adicionarlo al estrato mas parecido

Independientemente para cada dominio de estudio, se seleccionaron secciones (conglomerados de votantes) dentro de cada estrato, con probabilidad proporcional al tamaño y con reemplazo. La medida de tamaño fue la lista nominal de la sección y la asignación del número de secciones por estrato fue proporcional al tamaño del mismo. Finalmente, la selección se hizo de manera sistemática, ordenando las secciones según el porcentaje obtenido por la primera fuerza.

c) Cálculo del tamaño de muestra

Puesto que cada municipio era dominio de estudio, el cálculo del tamaño de muestra se hizo a nivel municipal. Para cada uno interesaba estimar proporciones[6], con una confianza del 95%, un error relativo máximo del 10% (10% del valor de P), un efecto del diseño de 26 y una tasa de no respuesta del 5%[7].

El número de votantes en muestra que se requiere para estimar la proporción P de interés se obtuvo mediante:

$$
\begin{aligned}
n_o &= \frac{t_{\alpha/2}^2 S^2}{d^2} Deff \\
&= \frac{t_{\alpha/2}^2 P(1-P)}{(rP)^2} Deff \\
&= \frac{1.96^2(1-P)}{(.10P)^2} 26.
\end{aligned}
$$

Efectuando la corrección por finitud:

$$
n = \frac{n_o}{1 + n_o / N}
$$

Para obtener el número de conglomerados, el valor anteior de n se dividió por el número promedio de votantes por sección. Finalmente, el número de secciones en muestra m, considerando la tasa de no respuesta, se obtuvo dividiendo el número de conglomerados por 0.95.

Lo anterior, aunado a consideraciones de costos y a la experiencia en proyectos anteriores, dio lugar a los siguientes tamaños de muestra por dominio de estudio:

---

6 Para cada dominio la(s) proporciónes) de interés correspondicron a la primera fuerza en el caso de que predominara un partido, y a las dos primeras fuerzas en el caso de municipios competidos

7 Los valores del efecto de diseño y la tasa de no respuesta se determinaron con base en levantamientos anteriores efectuados por la misma empresa. Sin embargo, dichos valores pueden variar entre proyectos y más aún, entre empresas, dependiendo de las circunstancias y supuestos que se hagan. Por ejemplo, la empresa "Berumen y Asociados" consideró un efecto de diseño general del orden de 46 en la planeación del conteo rápido para las elecciones a nivel nacional de 1994.

Cuadro 2.3.2.3
Tamaños de muestra por municipio

MUNICIPIO	*NUMERO DE SECCIONES EN LA POBLACION*	*NUMERO DE SECCIONES EN MUESTRA PLANEADAS*	*NUMERO DE SECCIONES EN MUESTRA DIFERENTES*[8]
Atlacomulco	39	27	25
Ixtlahuaca	48	25	25
Jilotepec	44	26	25
Nezahualcóyotl	660	40	40
Tlalnepantla.	362	40	40
Toluca	241	41	41
Valle de Chalco	103	40	40

Aunque en la práctica no todas las secciones llegan antes de la hora en que la empresa se compromete a dar los resultados, se espera que el porcentaje de secciones faltantes a esa hora sea pequeño debido al tipo de municipios objeto de estudio, además de que previamente se ha considerado una tasa de no respuesta.

d) Operativo de Campo

El trabajo de campo se realizó mediante un equipo integrado por coordinadores, supervisores y encuestadores. Para su mejor desempeño, días previos al levantamiento se les proporcionó un manual, asimismo, se impartió un curso de capacitación en el que se expusieron los objetivos del proyecto, las funciones y responsabilidades del personal, los formatos que deberían de llenar e instrucciones de transmisión al centro de recolección de información (centro de acopio).

Otra labor importante, una vez obtenida la muestra, consistió en localizar cada una de las secciones elegidas, de tal manera que se tuviera su ubicación, las principales vías de acceso así como medios de comunicación, y en su caso tiempos estimados para llegar al lugar desde el cual pudieran transmitir los resultados de la sección a su cargo.

Finalmente, el día de la elección, el personal indicado recolectó los resultados publicados a las afueras de las casillas y los transmitió al centro de acopio para su procesamiento.

e) Recepción y procesamiento de la información

Fue necesario organizar un centro de acopio, en donde se capturó la información de manera simultánea a su transmisión por parte del personal de campo. Para ello, se contó con un sistema integral compuesto por tres módulos:

[8] Ya que la selección fue con reemplazo, en el municipio de Atlacomulco dos secciones se eligieron 2 veces y en el municipio de Jilotepec una sección fue seleccionada 2 veces

- Captura, y en su caso, edición de la información
- Procesamiento de la misma
- Presentación de resultados

El primer módulo contempló algunos criterios para verificar la validez de la información capturada, así como garantizar que en la base de datos sólo se incluyeran las secciones en la muestra. El módulo de procesamiento obedeció al diseño de muestreo y los estimadores utilizados. Por último, mediante el tercer módulo, se presentaron numérica y gráficamente las pantallas de resultados una vez procesada la información.

2.3.3. Estimaciones

Dado el esquema de muestreo utilizado, el estimador[9] de la proporción de votos para cada partido estuvo dado por:

$$p(j)_{ppt} = \frac{y(j)_{ppt}}{N_{m.ppt}} = \frac{\sum_{h=1}^{H} y(j)_h}{\sum_{h=1}^{H} N_{m_h}} = \frac{\sum_{h=1}^{H} \frac{1}{m_h} \sum_{i=1}^{m_h} Y(j)_{hi} / \pi_{hi}}{\sum_{h=1}^{H} \frac{1}{m_h} \sum_{i=1}^{m_h} N_{hi} / \pi_{hi}}$$

donde:

$y(j)_h$ es el estimador insesgado del número total de votos para el partido j en el estrato h, $y(j)_{ppt} = \sum_{h=1}^{H} y(j)_h$

N_{m_h} es el estimador insesgado del número total de votos en el estrato h, $N_{m.ppt} = \sum_{h=1}^{H} N_{m_h}$

$p(j)_{ppt}$ es la proporción estimada de votos para el partido j

$Y(j)_{hi}$ es el número de votos para el partido j en la sección i del estrato h

N_{hi} es el número de votos en la sección i del estrato h

π_{hi} es la probabilidad de selección de la sección i dentro del estrato h, $\pi_{hi} = \frac{lista_{hi}}{lista_h}$

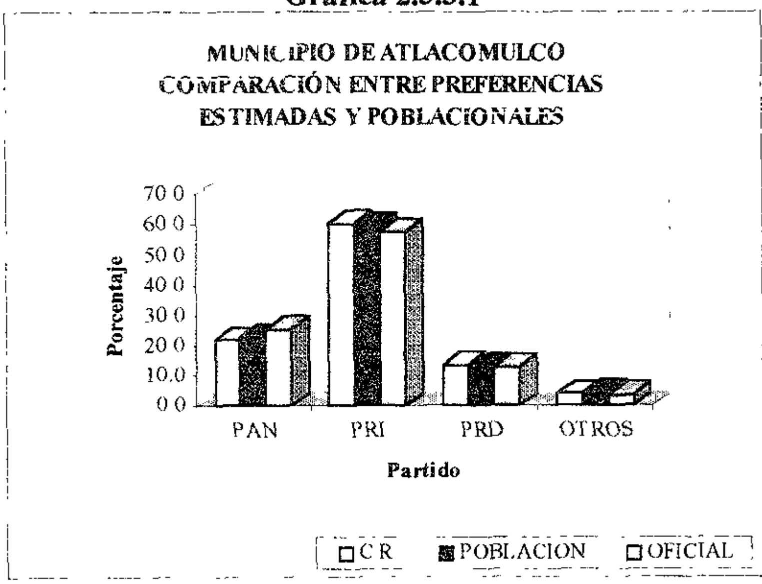
m_h es el número de secciones en muestra del estrato h

H es el número de estratos

[9] Hansen, Morris H. et. al. *Sample Survey Methods and Theory Vol. I.* John Willey & Sons. New York, 1953. p. 316.

Luego, el porcentaje estimado de votos para el partido j se obtuvo multiplicando la proporción respectiva por 100. En las gráficas 2.3.3.1 a 2.3.3.7 se muestran los resultados estimados de acuerdo con la fórmula anterior, así como los resultados poblacionales[10].

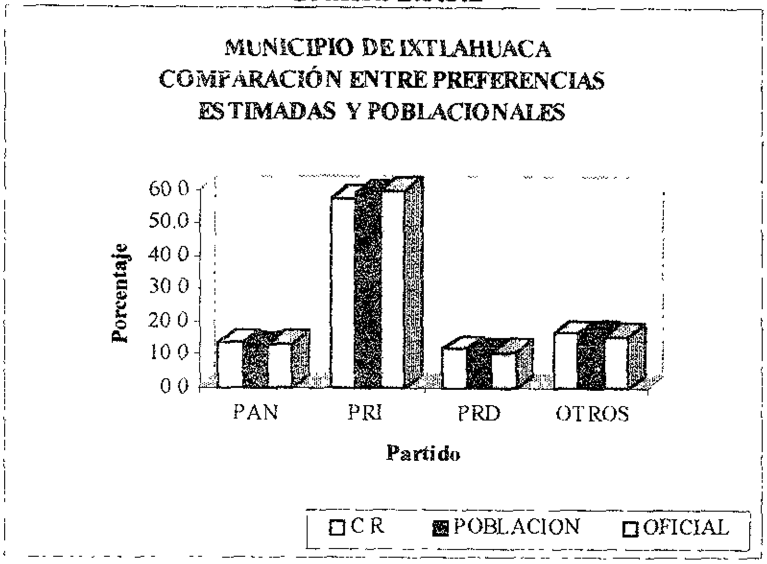
Gráfica 2.3.3.1



PARTIDO	*CONTEO RAPIDO*	*POBLACIONAL*	*OFICIAL*[11]
PAN	21.8	22.8	25.3
PRI	60.5	59.9	57.8
PRD	13.6	13.0	13
OTROS	4.2	4.3	3.9

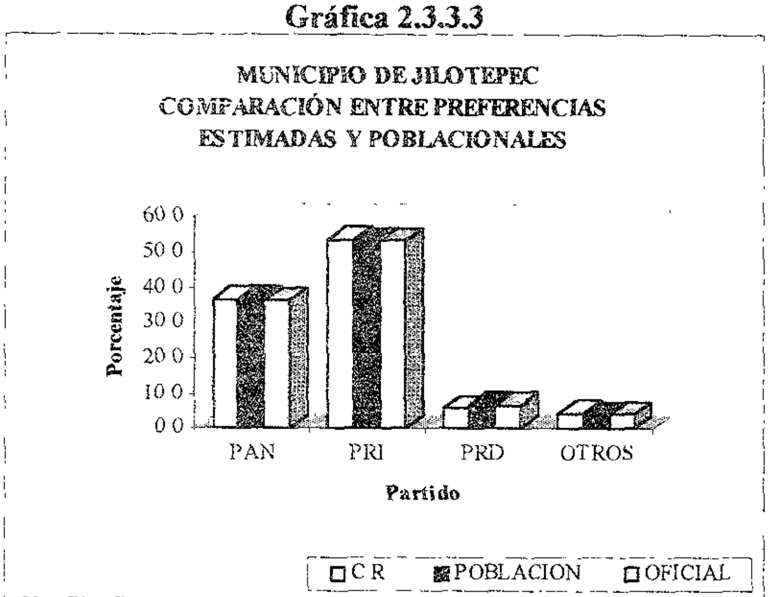
[10] Estos resultados fueron calculados con la base de datos proporcionada por el IFE, y difieren de los oficiales debido a las impugnaciones o bien a las casillas que no se hubieran recibido al momento en que el IFE proporcionó la información por casilla.
[11] Fuente: Resultados definitivos IFE

Gráfica 2.3.3.2



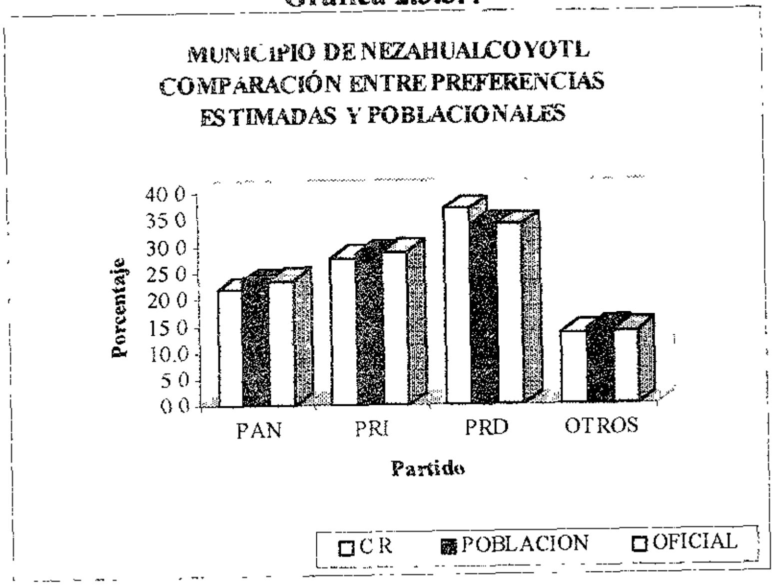
PARTIDO	CONTEO RAPIDO	POBLACIONAL	OFICIAL
PAN	13.8	13.2	13.3
PRI	57.5	59.0	59.8
PRD	11.9	11.0	11.1
OTROS	16.9	16.8	15.8

Gráfica 2.3.3.3



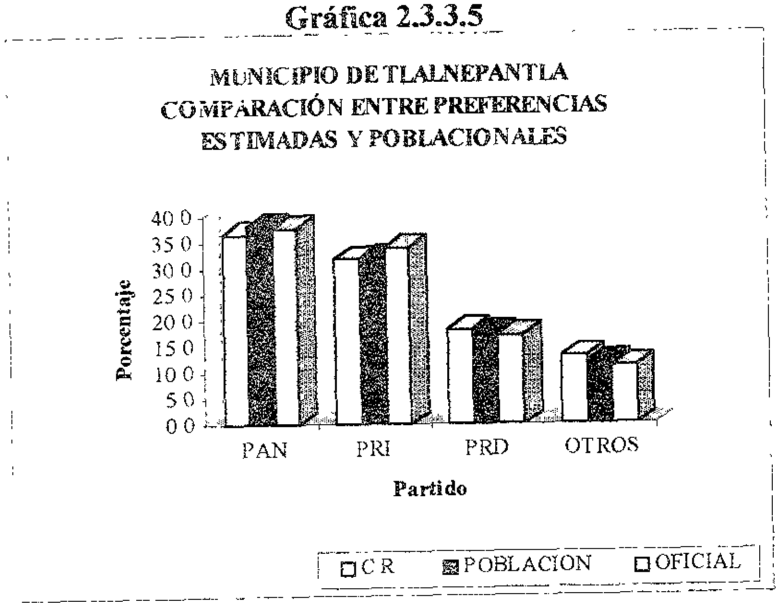
PARTIDO	CONTEO RAPIDO	POBLACIONAL	OFICIAL
PAN	36.3	36.3	36.2
PRI	53.5	53.0	53.0
PRD	5.8	6.6	6.7
OTROS	4.3	4.1	4.1

Gráfica 2.3.3.4



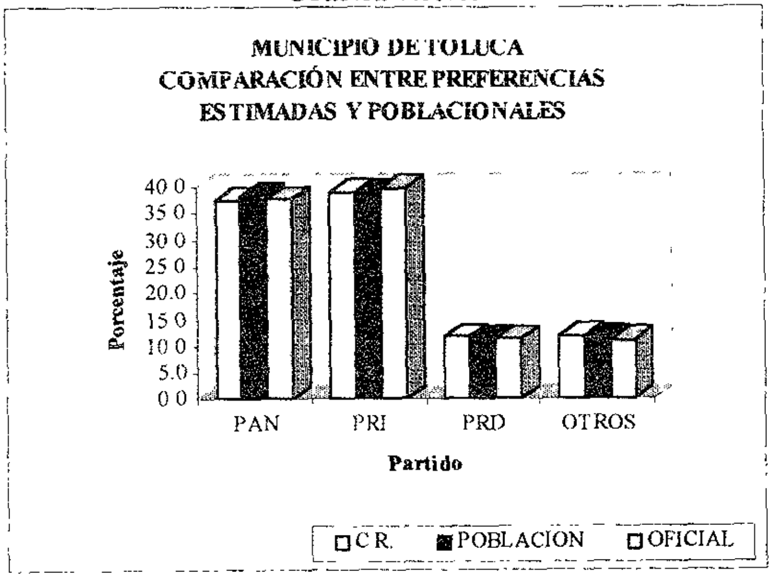
PARTIDO	*CONTEO RAPIDO*	*POBLACIONAL*	*OFICIAL*
PAN	21.8	23.3	23.5
PRI	27.8	28.6	28.8
PRD	36.9	33.8	34.0
OTROS	13.5	14.3	13.7

Gráfica 2.3.3.5



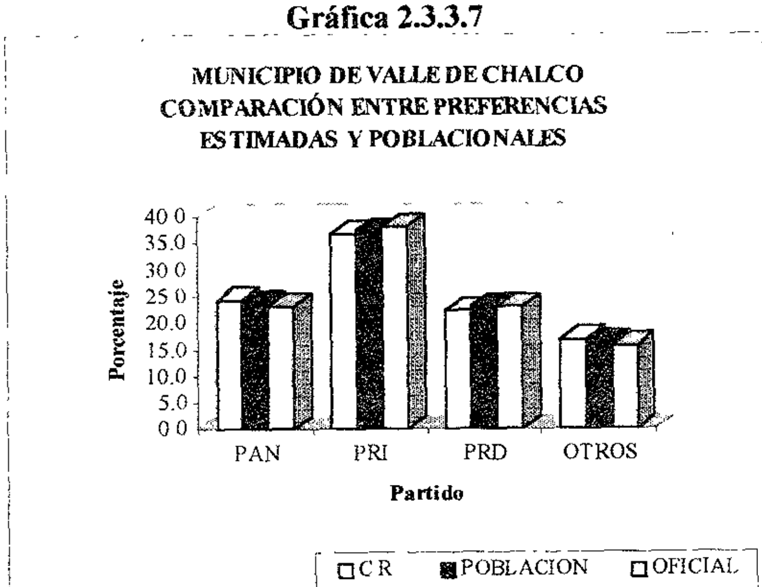
PARTIDO	*CONTEO RAPIDO*	*POBLACIONAL*	*OFICIAL*
PAN	36.8	38.3	37.7
PRI	31.9	32.2	33.9
PRD	18.3	17.5	17.1
OTROS	13.0	12.0	11.3

Gráfica 2.3.3.6



PARTIDO	*CONTEO RAPIDO*	*POBLACIONAL*	*OFICIAL*
PAN	37.5	38.2	37.7
PRI	38.9	39.0	39.7
PRD	11.8	11.3	11.4
OTROS	11.8	11.5	11.2

Gráfica 2.3.3.7



PARTIDO	*CONTEO RAPIDO*	*POBLACIONAL*	*OFICIAL*
PAN	24.2	23.7	23.1
PRI	36.7	37.3	38.1
PRD	22.3	23.0	23.0
OTROS	16.7	16.0	15.7

Como se puede apreciar, las estimaciones fueron muy cercanas al parámetro. En principio esto nos sugiere que el esquema de muestreo, el tamaño de muestra y el estimador utilizado fueron adecuados. No obstante, surge una inquietud: ¿para el mismo número de secciones, existe otro u otros esquemas de muestreo, tal vez más sencillos, que tengan una precisión similar?. Si la respuesta es afirmativa podríamos obtener estimadores más sencillos. En caso contrario, habría que evaluar si la pérdida en precisión se compensa con la facilidad de cálculo tanto del estimador como de su varianza.

3. ANÁLISIS Y PRESENTACIÓN DE RESULTADOS

Puesto que la encuesta corresponde a un conteo rápido, las secciones representan de manera natural conglomerados de diferente tamaño. En teoría, cuando los conglomerados varían en tamaño pueden originar que la varianza de las estimaciones sea mayor. Sin embargo, existen diferentes formas de incrementar la precisión del estimador, entre las que se encuentran:

- Utilizar estimadores de razón, que a pesar de ser sesgados, son consistentes, es decir, el sesgo disminuye conforme el número de unidades primarias de muestreo se incrementa.
- Utilizar un muestreo con probabilidad proporcional al tamaño (ppt).
- Estratificar de acuerdo al tamaño del conglomerado[1].

Si bien teóricamente existe una "ganancia" por utilizar un muestreo con probabilidad proporcional al tamaño, es de interés práctico saber qué tan grande es esta ganancia para el caso de algunos conteos rápidos. Con esta finalidad primero se comparan los siguientes esquemas de muestreo:

a) Muestreo estratificado con selección de conglomerados con igual probabilidad y con reemplazo.
b) Muestreo estratificado con selección de conglomerados con igual probabilidad y sin reemplazo.
c) Muestreo estratificado con selección de conglomerados con probabilidad proporcional al tamaño y con reemplazo.

En los tres casos la estratificación fue la misma y se basó en información sobre comicios anteriores. Con la finalidad de ver si se obtuvo alguna ganancia con la estratificación, se analizaron los diseños que se señalan a continuación:

d) Muestreo con selección de conglomerados con igual probabilidad y con reemplazo.
e) Muestreo con selección de conglomerados con igual probabilidad y sin reemplazo (muestreo aleatorio de conglomerados).
f) Muestreo con selección de conglomerados con probabilidad proporcional al tamaño y con reemplazo.

Uno de los mejores estimadores cuando se tienen conglomerados de tamaño variable es el estimador de razón[2], por lo que en los tres diseños se consideraron estimadores de este tipo.

---

[1] En la comparación no se estratificó por este criterio ya que se poseía información sobre otra variable, y se prefirió controlar el sesgo debido a la variación en el tamaño mediante los estimadores de razón y el muestreo con ppt.

[2] Existen otros estimadores como el "estimador medio", que es un promedio de las medias de los conglomerados en muestra, y el "estimador insesgado", que también es un promedio de medias, pero cada media del conglomerado se calcula con base al tamaño promedio de los conglomerados.
Ver Sukhatme, P.V. *et. al. Sampling Theory of Surveys with Applications*. 3ª edición, Iowa State University Press. Iowa, 1984, p. 290.

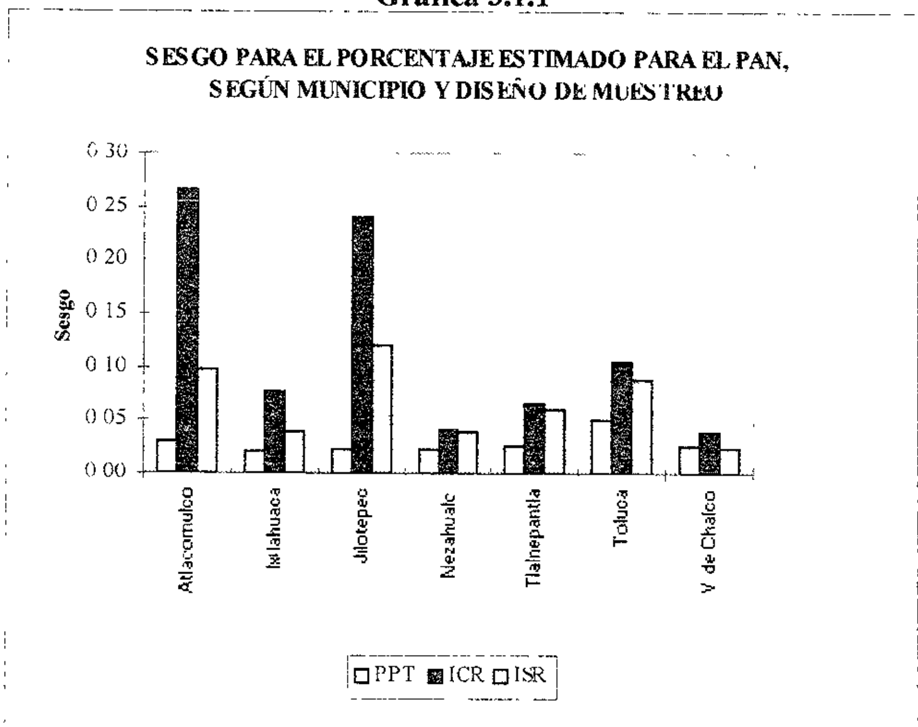
La comparación se hizo principalmente en términos del efecto de diseño con respecto a un m.a.s.s.r., y de la eficiencia relativa de los diseños. Las razones por las cuales se utilizó el efecto de diseño fueron varias: involucra la varianza del estimador, su posible utilización en conteos rápidos posteriores y se interpreta de forma sencilla. A pesar de ello, también tiene limitaciones, como se verá más adelante, lo cual implica que se debe ser cuidadoso al utilizarlo. Por ello, se analizaron también las varianzas y sesgos de los estimadores, así como coeficientes de variación.

Ya que se contó con información de toda la población, es decir, de los resultados de todas las secciones del municipio y no sólo de las secciones en muestra, no fue necesario estimar las varianzas de los estimadores, sino que se utilizaron las fórmulas que aparecen en el Anexo B. En lo que se refiere al número de secciones en muestra, se mantuvo fijo para cada municipio y correspondieron a los que se mencionaron en el inciso 2.3.2.

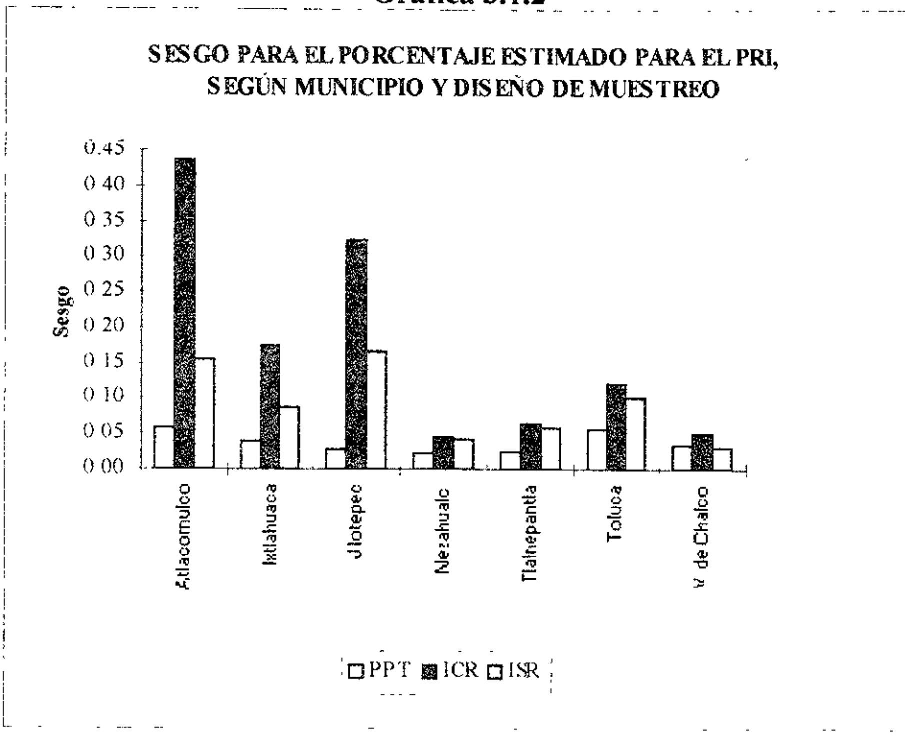
3.1. Sesgo

En cuanto al sesgo para el porcentaje estimado, este resultó ser muy pequeño para los tres diseños (estratificados), siendo menor a un punto porcentual (gráficas 3.1.1 a 3.1.3). Si bien el sesgo fue muy pequeño, el menor corresponde al esquema con ppt, le sigue el muestreo con igual probabilidad y sin reemplazo (ISR) y, por último, el esquema con igual probabilidad y reemplazo (ICR). Así, la selección con ppt disminuyó el tamaño del sesgo.

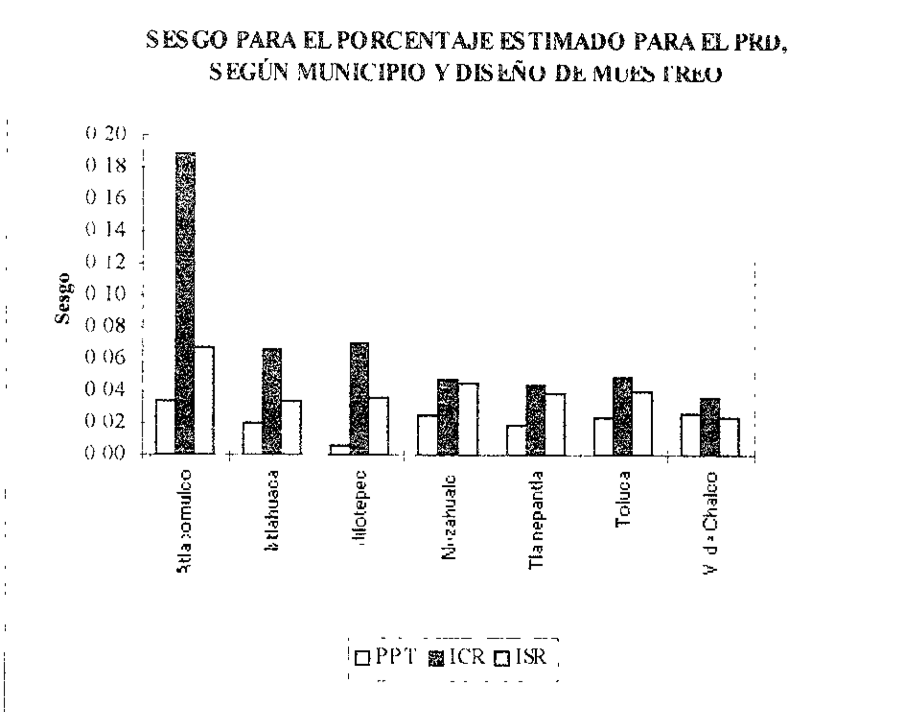
Gráfica 3.1.1



Gráfica 3.1.2



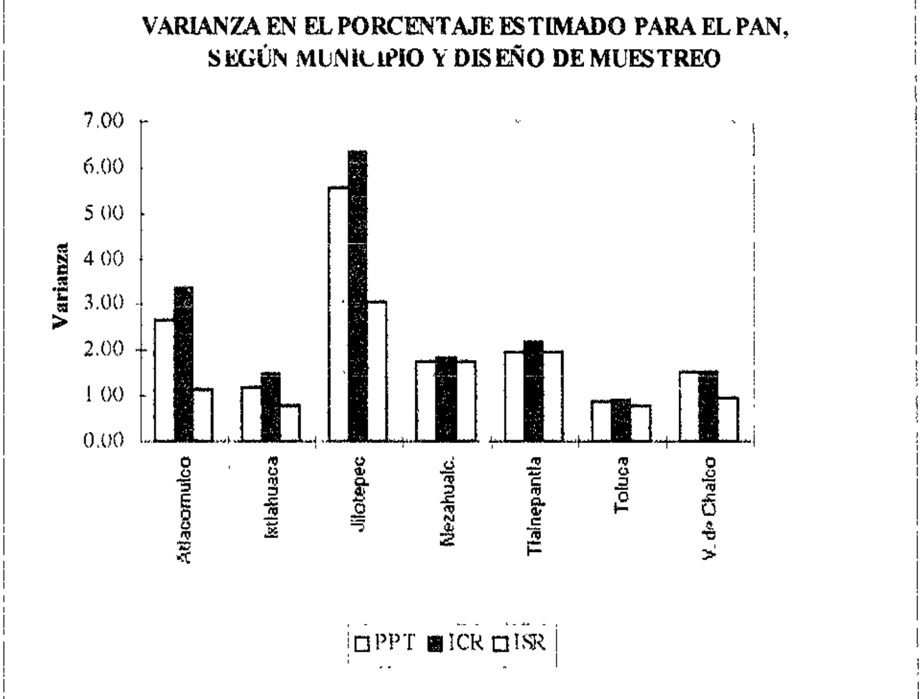
Gráfica 3.1.3



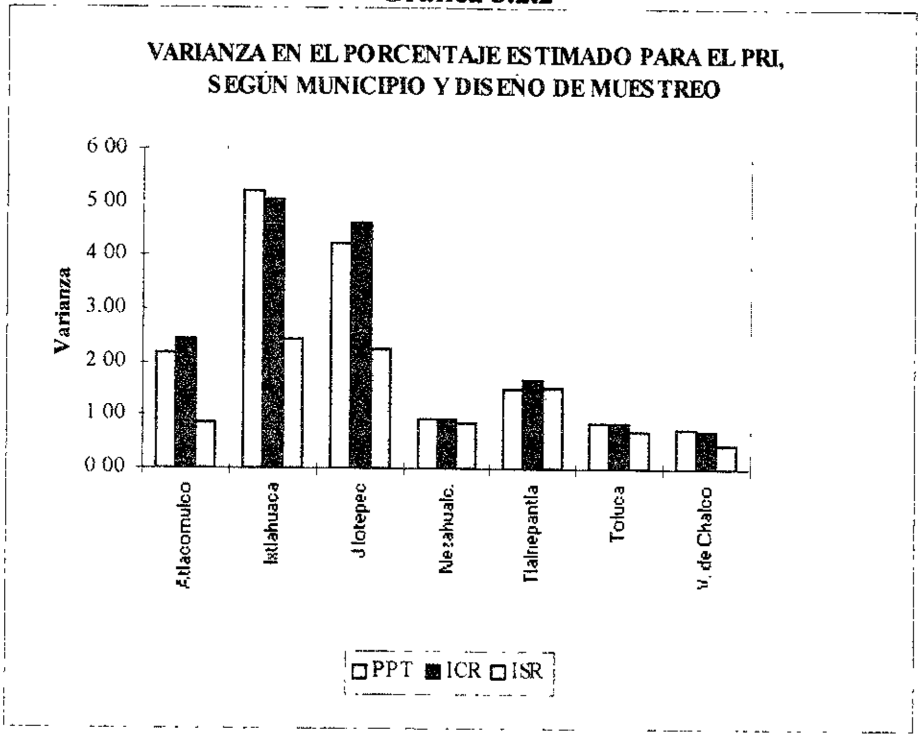
3.2. Varianzas

En las gráficas 3.2.1 a 3.2.3 se muestra la varianza del porcentaje estimado para cada partido y dominio de estudio. Como puede observarse, en general, la varianza del muestreo ppt (con reemplazo) fue ligeramente más pequeña o igual que la del diseño con igual probabilidad y reemplazo, mientras que fue mayor a la del muestreo con igual probabilidad y sin reemplazo.

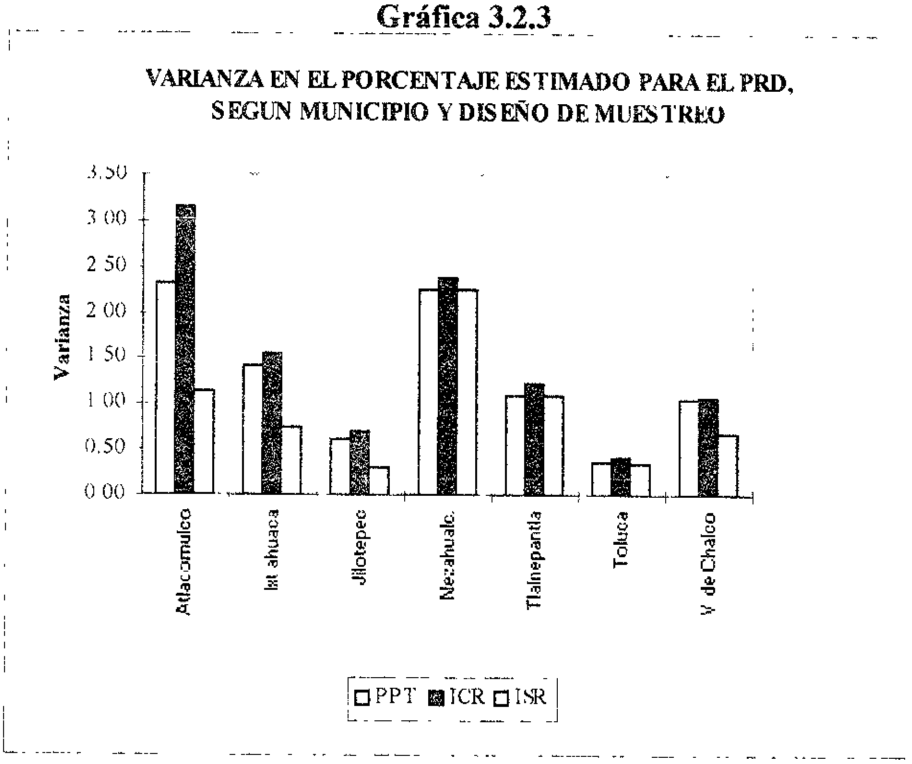
Gráfica 3.2.1



Gráfica 3.2.2

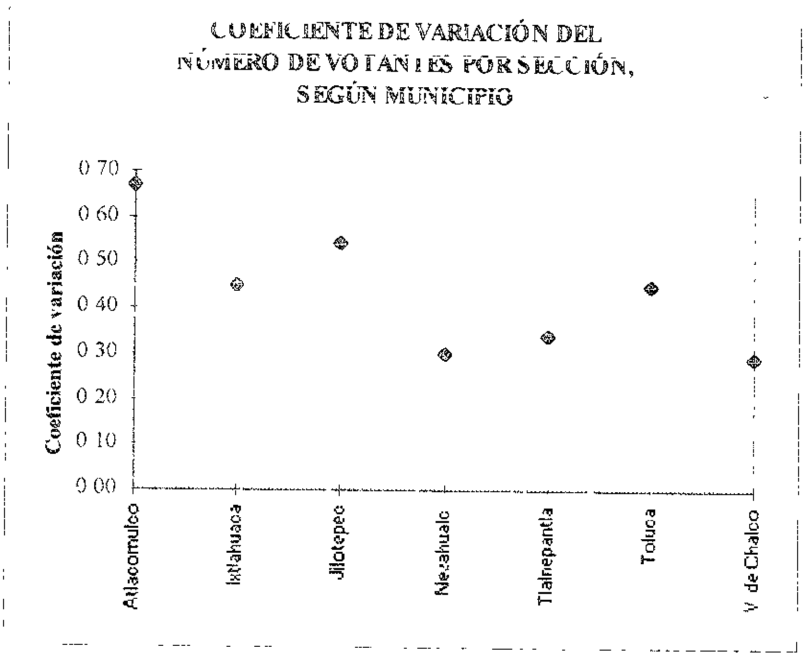


Gráfica 3.2.3



El que se tenga una ganancia pequeña del muestreo con probabilidades diferentes con respeto al de probabilidades iguales, ambos con reemplazo, se explica en parte porque cuando se estiman promedios o porcentajes, una variación moderada en el tamaño del conglomerado tiene un efecto pequeño en el aumento de precisión. Al respecto, en la literatura[3] se señala que esto es cierto si el coeficiente de variación del tamaño del conglomerado es menor a 0.5 o 1, como ocurrió para los siete municipios (ver gráfica 3.2.4).

Gráfica 3.2.4



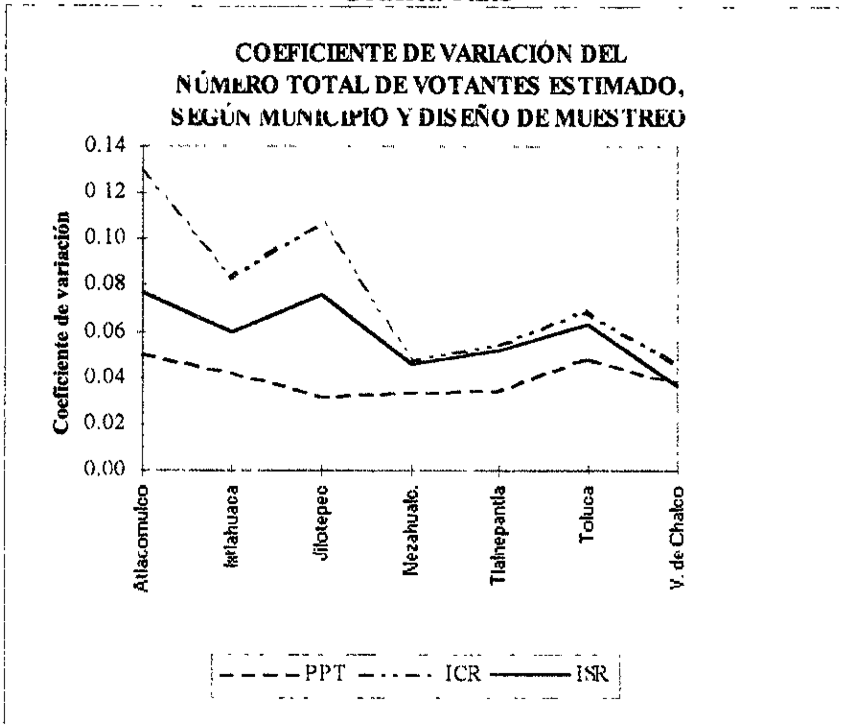
Así, en el caso de porcentajes y promedios, se deberá hacer un balance entre la dificultad en reducir la variación en el tamaño y la ganancia en precisión, ya que esta ganancia suele ser moderada.

Por otra parte, la aproximación para la varianza del estimador de razón supone que el coeficiente de variación del número total de votantes estimado, denominador del cociente, es pequeño. De hecho, algunos textos[4] señalan que para que la aproximación sea buena, debe cumplirse que tal coeficiente de variación sea menor a 0.10. Como se aprecia en la gráfica 3.2.5 el único municipio que estuvo ligeramente arriba de este valor fue Atlacomulco.

---

[3] Hansen, Morris H et. al. *Sample Survey Methods and Theory Vol. I.* John Willey & Sons New York, 1953. p. 354.
[4] Kish, Leslie *Muestreo de Encuestas* 1ª edición Trillas México, 1792 p 225

Gráfica 3.2.5



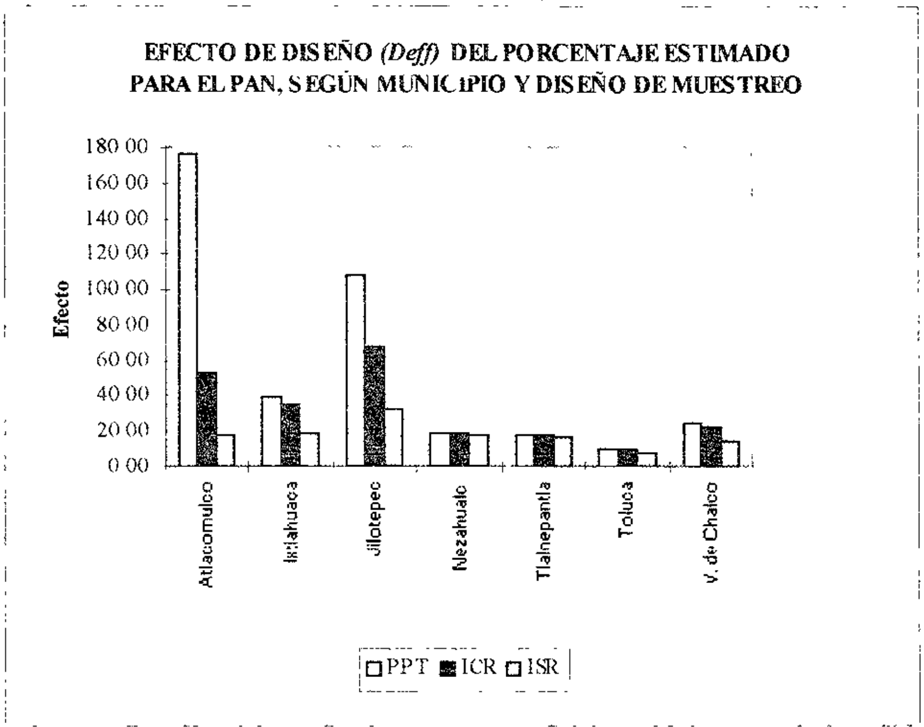
3.3. Efectos de diseño

La situación de los efectos de diseño requiere de mayor profundidad. Primero se consideran los efectos de diseño del esquema complejo con respecto a si se tuviera un muestreo aleatorio de votantes, y posteriormente, se analizan los efectos del diseño correspondiente tomando como referencia un muestreo aleatorio simple de conglomerados.

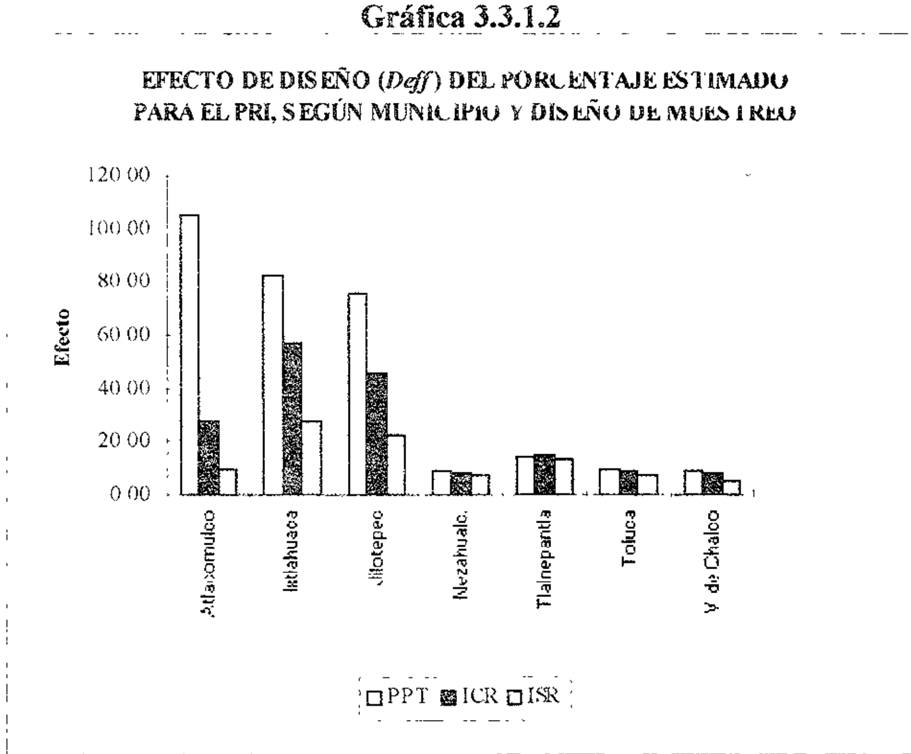
3. 3. 1. Efectos de diseño con respecto a un muestreo aleatorio de votantes

En lo referente al efecto de diseño con respecto a si se tuviera un muestreo aleatorio de votantes (*Deff*) , resultó ser mayor para el muestreo con probabilidad proporcional con respecto a los otros dos diseños, sobretodo en los municipios pequeños de Atlacomulco, Ixtlahuaca y Jilotepec. En los municipios restantes, los efectos de diseño fueron parecidos, siendo ligeramente mayores en el muestreo con ppt. Para los siete municipios fue menor el del esquema con igual probabilidad y sin reemplazo (gráficas 3.3.1.1 a 3.3.1.3).

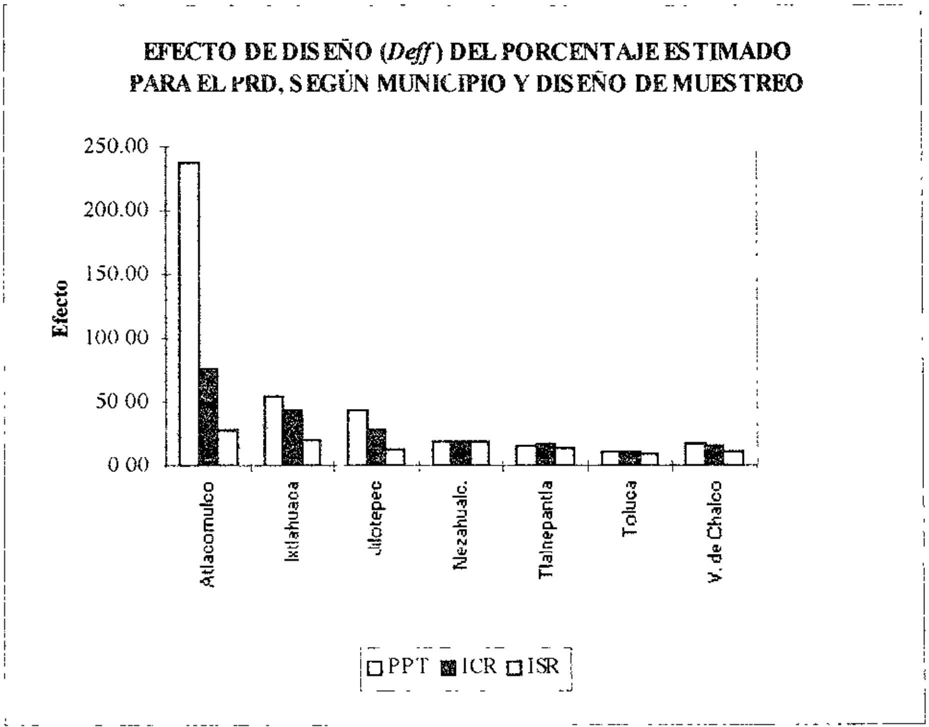
Gráfica 3.3.1.1



Gráfica 3.3.1.2



Gráfica 3.3.1.3



Las razones por las que los valores de los efectos de diseño para el ppt, en algunos casos, fueron muy altos, y en general, mayores a las de los muestreos con igual probabilidad pueden ser: la fracción de muestreo y el número esperado de votantes en muestra.

En lo que se refiere a la primera, hay que tomar en cuenta que el muestreo con igual probabilidad y sin reemplazo se ve favorecido cuando la fracción de muestreo es grande, como en el caso de los municipios de Atlacomulco, Ixtlahuaca , Jilotepec y Valle de Chalco

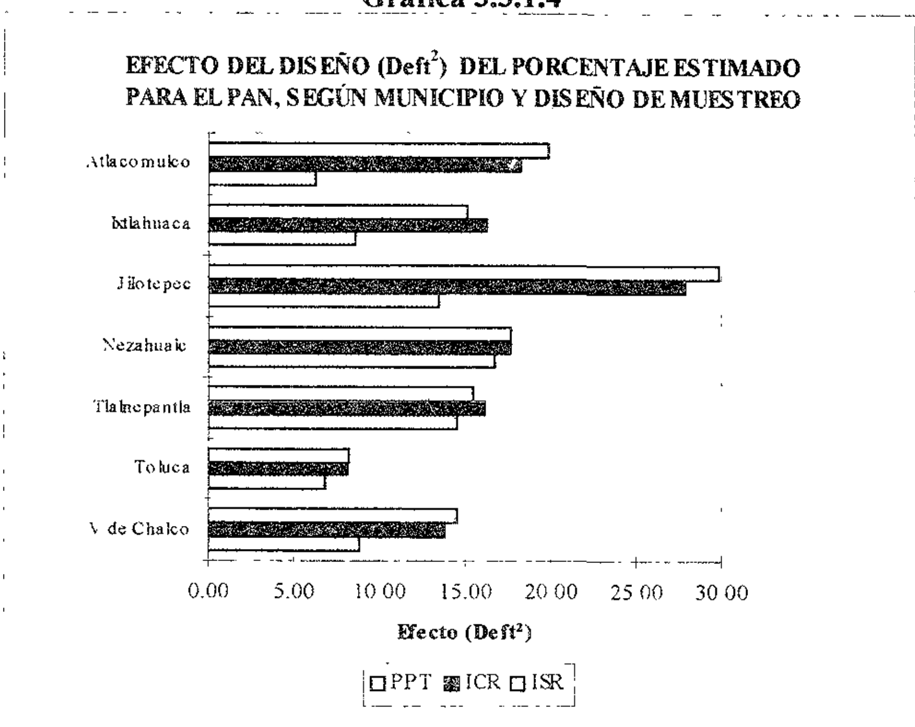
Cuadro 3.3.1.1
Número de secciones, según municipio

MUNICIPIO	*MUESTRA*	*POBLACION*	*FRACCION DE MUESTREO (ISR)*
Atlacomulco	25	39	0.64
Ixtlahuaca	25	48	0.52
Jilotepec	25	44	0.57
Nezahualcóyotl	40	660	0.06
Tlalnepantla	40	362	0.11
Toluca	41	241	0.17
Valle de Chalco	40	103	0.39

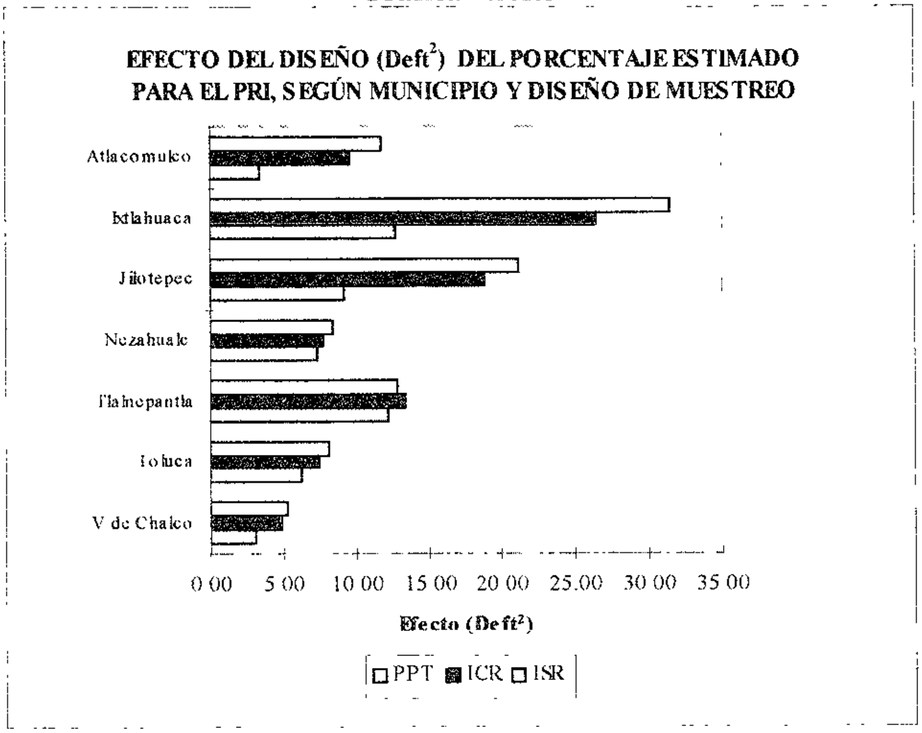
Debido a que la fracción de muestreo era muy alta para algunos municipios, se recurrió a la segunda definición del efecto *Deft*^{2} (apartado 1.5), con la finalidad de controlar la influencia de la misma en el efecto de diseño. Es decir, se comparó la varianza del diseño respectivo con la de un m.a.s.c.r de votantes.

En las gráficas 3.3.1.4 a 3.3.1.6 se observa que los efectos de diseño para el muestreo con ppt, y los del muestreo con igual probabilidad y reemplazo fueron muy similares. Al igual que en el caso del *Deff*, el *Deft*^{2} para el esquema con igual probabilidad y sin reemplazo resultó ser el menor. Se aprecia también que la variación en los efectos de diseño con respecto a un m.a.s.c.r. fue menor que la variación en los efectos con respecto a un m.a.s.s.r.

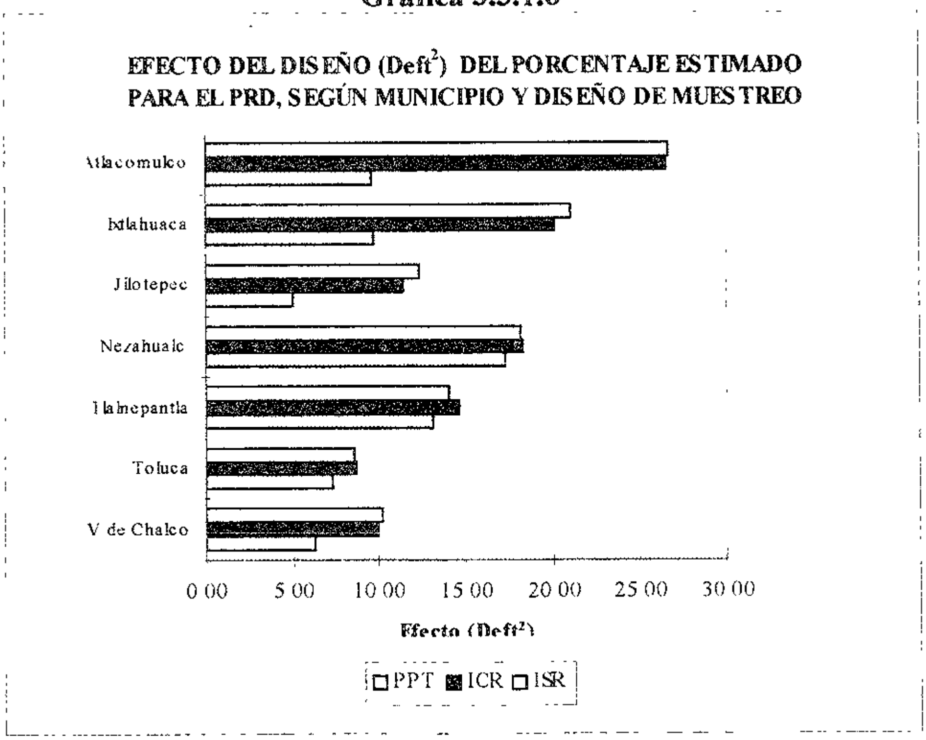
Gráfica 3.3.1.4



Gráfica 3.3.1.5



Gráfica 3.3.1.6



La segunda razón a considerar fue el número esperado de votantes (ver anexo B), el cual bajo un diseño proporcional al tamaño es mayor que bajo un diseño con igual probabilidad. Esto dio lugar a que la varianza para el muestreo irrestricto aleatorio fuera menor en el primer caso y por ende el efecto de diseño mayor.

Cuadro 3.3.1.2
Número esperado de votantes en muestra por municipio, para un número fijo de conglomerados (secciones)

MUNICIPIO	*PROBABILIDAD PROPORCIONAL*	*CON IGUAL PROBABILIDAD*	*NUM. DE SECCIONES EN MUESTRA*
Atlacomulco	12,993	9,551	25
Ixtlahuaca	14,603	12,658	25
Jilotepec	12,472	10,156	25
Nezahualcóyotl	17,991	17,180	40
Tlalnepantla	18,653	17,440	40
Toluca	22,896	20,904	41
Valle de Chalco	17,024	16,515	40

3. 3. 2. Efectos de diseño con respecto a un muestreo aleatorio de secciones

Puesto que el efecto de diseño es una medida de la eficiencia relativa de un m.a.s. con respecto a otro esquema, y que el marco de muestreo fueron secciones, otra medida importante es el efecto de diseño de cada uno de los tres diseños comparados hasta el momento, con respecto a un muestreo aleatorio simple de conglomerados[5], denotado por $Deff_c$.

En el cuadro 3.3.2.1 se observa que en los municipios de Atlacomulco, Ixtlahuaca, Jilotepec y Valle de Chalco, municipios en los que predomina un partido y cuya fracción de muestreo fue alta, el efecto de diseño para el muestreo con igual probabilidad y reemplazo fue un poco más grande o similar que el efecto del muestreo con probabilidad proporcional al tamaño. El efecto más pequeño correspondió al muestreo con igual probabilidad sin reemplazo.

Con respecto a los municipios en los que la fracción de muestreo fue pequeña y además fueron competidos electoralmente, Tlalnepantla, Nezahualcóyotl y Toluca, el efecto del diseño fue muy parecido para los tres esquemas de muestreo.

[5] Muestreo aleatorio de conglomerados con igual probabilidad y sin reemplazo.

Cuadro 3.3.2.1
Efecto de diseño (Deff$_c$) con respecto al muestreo aleatorio de conglomerados, según partido y municipio

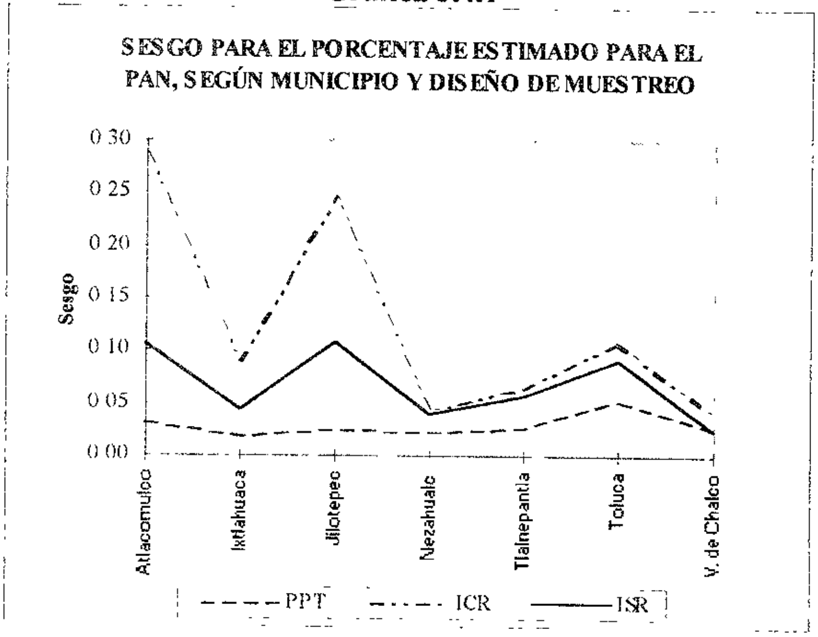
MUNICIPIO	*NUM. SECCIONES EN MUESTRA*	*PAN*			*PRI*			*PRD*		
		PPT	*ICR*	*ISR*	*PPT*	*ICR*	*ISR*	*PPT*	*ICR*	*ISR*
Atlacomulco	25	1.8	2.2	0.8	1.5	1.7	0.6	1.8	2.4	0.9
Ixtlahuaca	25	1.5	1.9	1.0	1.4	1.3	0.6	1.8	2.0	0.9
Jilotepec	25	2.0	2.3	1.1	1.9	2.0	1.0	1.4	1.6	0.7
Nezahualcóyotl	40	1.0	1.0	1.0	0.9	0.9	0.8	1.0	1.0	1.0
Tlalnepantla	40	0.6	0.7	0.6	0.7	0.8	0.7	0.7	0.8	0.7
Toluca	41	1.1	1.2	1.0	1.2	1.2	1.0	1.0	1.1	0.9
Valle de Chalco	40	1.6	1.6	1.0	1.6	1.5	1.0	1.6	1.6	1.0

3.4. Comparación entre muestreos estratificados y no estratificados.

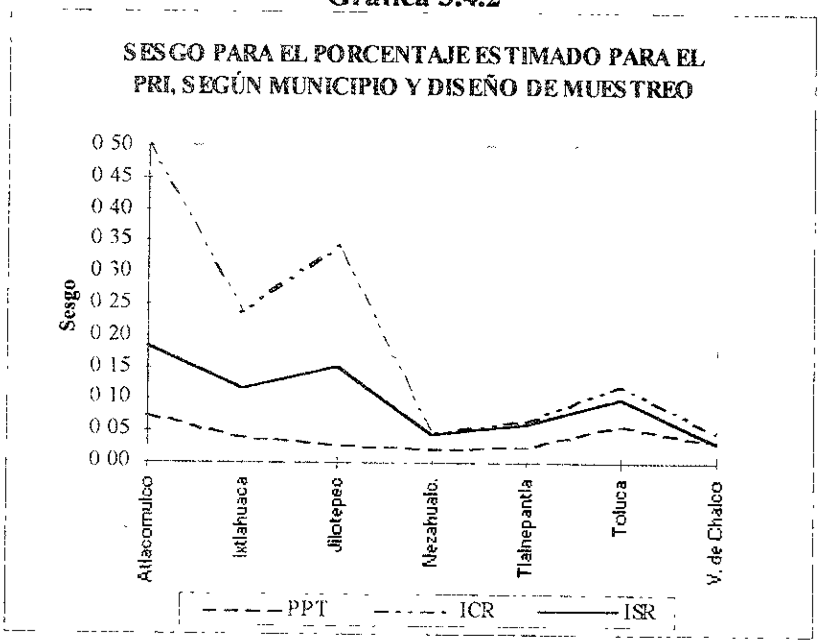
Hasta el momento principalmente se han hecho comparaciones entre esquemas que involucran una estratificación. Sin embargo, es de utilidad explorar qué tanto se gana con esta última. Para ello, primero se compararon los tres esquemas sin estratificar (proporcional al tamaño, igual probabilidad con reemplazo e igual probabilidad sin reemplazo) y posteriormente se compararon respecto a los estratificados.

Los resultados que se obtuvieron fueron análogos a cuando se hizo la estratificación. Así, en cuanto al sesgo, el menor corresponde al muestreo ppt, le sigue el muestreo aleatorio de conglomerados y el mayor fue para el esquema con igual probabilidad y reemplazo (ver gráficas 3.4 1 a 3.4.3).

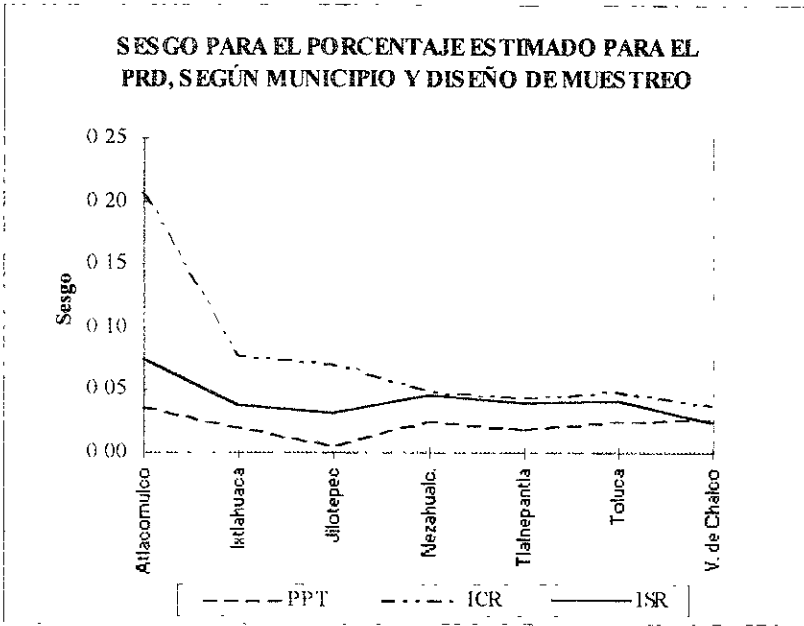
Gráfica 3.4.1



Gráfica 3.4.2

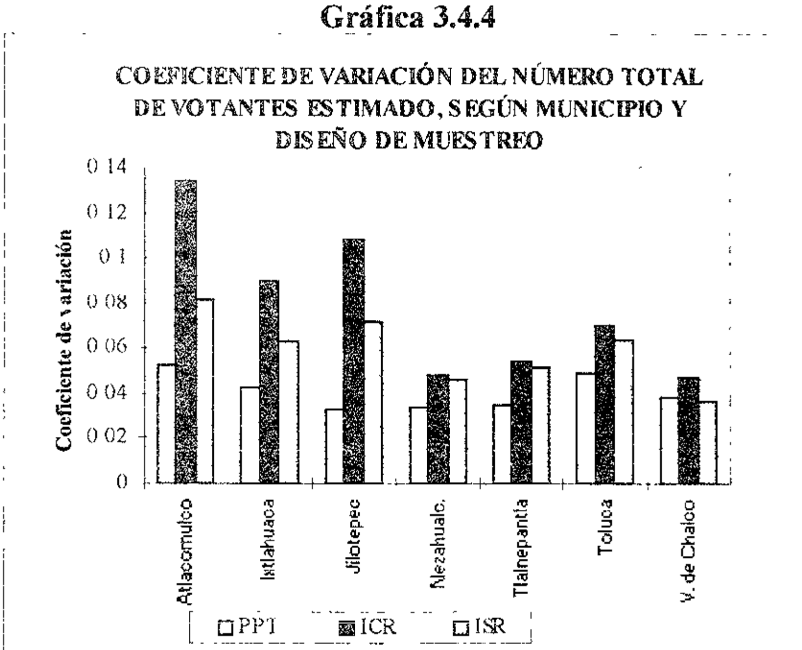


Gráfica 3.4.3



En lo que se refiere al coeficiente de variación del estimador del número total de votos, este fue menor para el esquema proporcional, y el mayor correspondió al de igual probabilidad y con reemplazo.

Gráfica 3.4.4



Al observar las eficiencias relativas resulta que en los municipios cuya fracción de muestreo fue grande, el diseño del muestreo sin reemplazo fue más eficiente que el ppt; en los municipios grandes como Nezahualcóyotl, Toluca y Tlalnepantla, la precisión fue similar bajo ambos tipos de muestreo. La varianza del muestreo con reemplazo e igual probabilidad fue muy parecida a la del ppt.

Cuadro 3.4.1.
Eficiencia relativa del muestreo con probabilidad proporcional respecto al de igual probabilidad con reemplazo, según partido y municipio

MUNICIPIO	*NUM. SECC. MUESTRA*	*PAN*	*PRI*	*PRD*
Atlacomulco	25	1.0	0.9	1.0
Ixtlahuaca	25	1 1	1.0	1 1
Jilotepec	25	1.0	1.0	1.0
Nezahualcóyotl	40	1.1	1.1	1.1
Tlalnepantla	40	1.3	1.1	1.3
Toluca	41	1.1	1.0	1.1
Valle de Chalco	40	1.1	1.0	1.1

Cuadro 3.4.2
Eficiencia relativa del muestreo con probabilidad proporcional respecto al de igual probabilidad y sin reemplazo, según partido y municipio

MUNICIPIO	*NUM. SECC. MUESTRA*	*PAN*	*PRI*	*PRD*
Atlacomulco	25	0.6	0.6	0.6
Ixtlahuaca	25	0.9	0.8	0.9
Jilotepec	25	1.0	0.9	1.0
Nezahualcóyotl	40	1.0	1.0	1.0
Tlalnepantla	40	0.5	0.4	0.5
Toluca	41	0.6	0.5	0.6
Valle de Chalco	40	0.5	0.5	0.5

Por otro lado, con la finalidad de ver si se obtuvo mayor precisión cuando se estratificó, se calculó la eficiencia relativa del diseño respectivo sin estratificar, con respecto al estratificado. Los resultados se muestran en los cuadros 3.4.3. a 3.4.5. Como se esperaba en todos los casos el muestreo estratificado fue igual o más eficiente que el muestreo sin estratificar. Por lo tanto, en términos de precisión existió una ganancia debido a la estratificación.

Cuadro 3.4.3.
Eficiencias relativas del muestreo estratificado con probabilidad proporcional respecto al muestreo sin estratificar con probabilidad proporcional, según partido y municipio

MUNICIPIO	PAN	PRI	PRD
Atlacomulco	1.2	1.6	1.1
Ixtlahuaca	1.2	1.5	1.0
Jilotepec	1.1	1.2	1.4
Nezahualcóyotl	1.0	1.2	1.0
Tlalnepantla	1.7	1.5	1.4
Toluca	1.0	1.0	1.1
Valle de Chalco	1.0	1.1	1.0

Cuadro 3.4.4
Eficiencias relativas del muestreo estratificado con igual probabilidad y con reemplazo respecto al muestreo sin estratificar con igual probabilidad y con reemplazo, según partido y municipio

MUNICIPIO	PAN	PRI	PRD
Atlacomulco	1.2	1.6	1.1
Ixtlahuaca	1.1	1.5	1.0
Jilotepec	1.0	1.1	1.4
Nezahualcóyotl	1.0	1.2	1.0
Tlalnepantla	1.7	1.4	1.4
Toluca	1.0	1.0	1.1
Valle de Chalco	1.0	1.1	1.0

Cuadro 3.4.5
Eficiencias relativas del muestreo estratificado con igual probabilidad y sin reemplazo respecto al muestreo sin estratificar con igual probabilidad y sin reemplazo, según partido y municipio

MUNICIPIO	PAN	PRI	PRD
Atlacomulco	1.3	1.7	1.1
Ixtlahuaca	1.0	1.6	1.1
Jilotepec	0.9	1.0	1.4
Nezahualcóyotl	1.0	1.2	1.0
Tlalnepantla	1.7	1.4	1.4
Toluca	1.0	1.0	1.1
Valle de Chalco	1.0	1.0	1.0

Lo anterior se observa también en términos de los efectos de diseño (*Deff*) (cuadros 3.4.6 al 3.4.8). De acuerdo con estos, en la mayoría de los casos los efectos del respectivo muestreo estratificado y no estratificado fueron iguales, o bien, el efecto del estratificado fue menor.

Cuadro 3.4.6

Efectos de diseño (Deff) para el PAN, según diseño de muestreo y municipio

MUNICIPIO	*% PARTIDO*	*NUM. SECC. MUES.*	*NUM. SECC. POBLA.*	*PPT*		*ICR*		*ISR*	
				Estratificado	*No Estratificado*	*Estratificado*	*No Estratificado*	*Estratificado*	*No Estratificado*
Atlacomulco	22.8	25	39	177.1	159.7	52.5	60.5	18.0	22.3
Ixtlahuaca	13.2	25	48	39.8	44.6	35.1	35.2	18.7	17.2
Jilotepec	36.3	25	44	107.5	113.1	67.6	62.3	32.8	27.5
Nezahualcóyotl	23.3	40	660	18.9	19.7	18.8	19.3	17.8	18.2
Tlalnepantla	38.3	40	362	17.6	30.0	18.2	30.3	16.3	27.0
Toluca	38.2	41	241	10.2	10.3	9.8	9.8	8.2	8.2
V. de Chalco	23.7	40	103	24.3	25.1	22.6	23.1	14.4	14.3

Cuadro 3.4.7

Efectos de diseño (Deff) para el PRI, según diseño de muestreo y municipio

MUNICIPIO	*% PARTIDO*	*NUM. SECC. MUES.*	*NUM. SECC. POBLA.*	*PPT*		*ICR*		*ISR*	
				Estratificado	*No Estratificado*	*Estratificado*	*No Estratificado*	*Estratificado*	*No Estratificado*
Atlacomulco	59.9	25	39	105.2	125.2	27.8	42.1	9.8	15.5
Ixtlahuaca	59.0	25	48	82.4	119.3	57.1	82.3	27.6	40.3
Jilotepec	53.0	25	44	75.9	88.7	45.7	46.8	22.3	20.7
Nezahualcóyotl	28.6	40	660	9.0	10.7	8.3	9.8	7.8	9.3
Tlalnepantla	32.2	40	362	14.6	21.2	15.2	21.2	13.7	18.9
Toluca	39.0	41	241	10.0	10.1	9.0	9.1	7.5	7.5
V. de Chalco	37.3	40	103	8.9	9.5	8.0	8.6	5 1	5.3

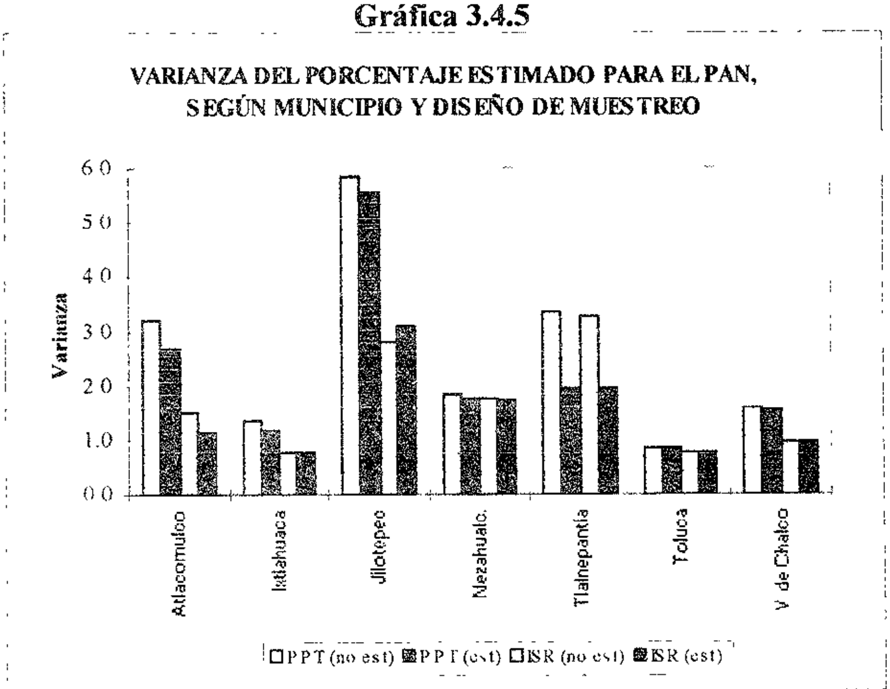
Cuadro 3.4.8

Efectos de diseño (Deff) para el PRD, según diseño de muestreo y municipio

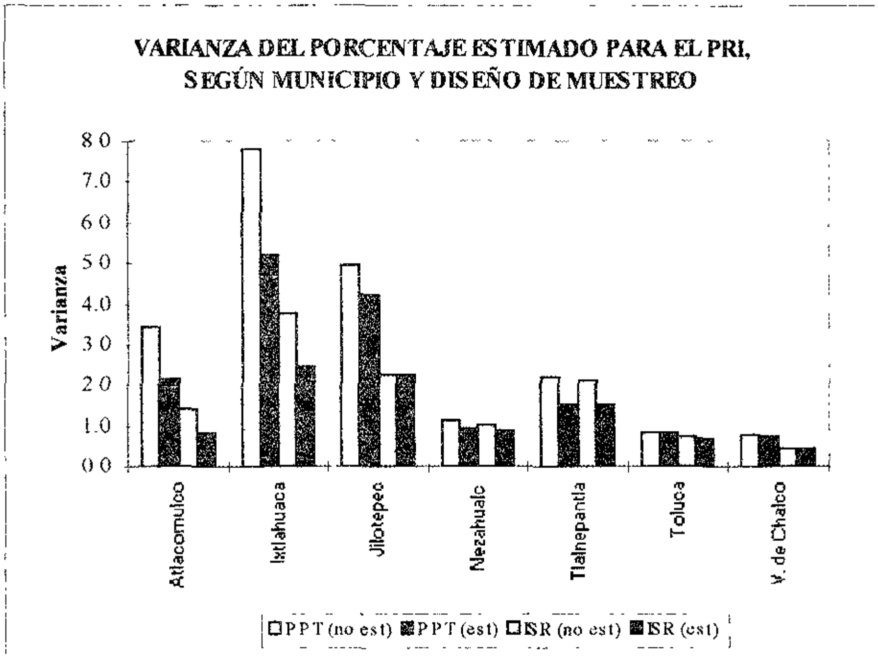
MUNICIPIO	*% PARTIDO*	*NUM. SECC. MUES.*	*NUM. SECC. POBLA.*	*PPT*		*ICR*		*ISR*	
				Estratificado	*No Estratificado*	*Estratificado*	*No Estratificado*	*Estratificado*	*No Estratificado*
Atlacomulco	13.0	25	39	237.9	199.3	76.4	80.4	27.7	29.6
Ixtlahuaca	11.0	25	48	55.3	54.5	43.4	42.7	21.0	20.9
Jilotepec	6.6	25	44	44.4	60.6	27.6	35.3	12.3	15.6
Nezahualcóyotl	33.8	40	660	19.4	20.4	19.5	20.2	18.5	19.0
Tlalnepantla	17.5	40	362	16.0	22.8	16.5	22.9	14.8	20.4
Toluca	11.3	41	241	10.6	11.7	10.5	11.4	8.8	9.5
V.de Chalco	23.0	40	103	16.9	17.5	16.3	16.7	10.3	10.3

Por último, ya que el muestreo con reemplazo e igual probabilidad es de poco interés práctico, a manera de resumen, se presentan las varianzas para los esquemas con probabilidad proporcional al tamaño (estratificado y sin estratificar) y las varianzas para los esquemas con igual probabilidad y sin reemplazo (estratificado y sin estratificar). En ellas se observa que en los municipios de Atlacomulco, Ixtlahuaca, Jilotepec y Valle de Chalco, el diseño con mayor precisión fue el estratificado de igual probabilidad y sin reemplazo. En Nezahualcóyotl y Toluca las varianzas fueron iguales, mientras que en Tlalnepantla, la precisión entre estratificados o bien no estratificados, fue muy similar.

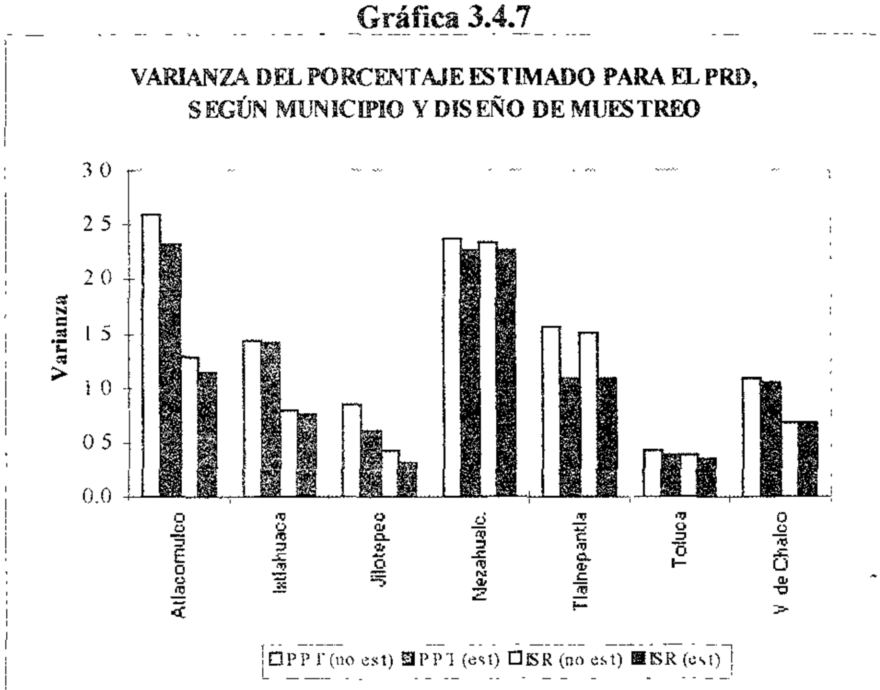
Gráfica 3.4.5



Gráfica 3.4.6



Gráfica 3.4.7



CONCLUSIONES

Las encuestas han tomado mayor importancia en épocas recientes, cada vez su uso es más frecuente y han sido aceptadas gradualmente como un medio para obtener información de manera alternativa a un censo, pero de manera más económica, rápida y, a veces, más confiable.

La elaboración de encuestas comprende diversas etapas relacionadas entre sí, una de ellas corresponde al esquema de muestreo que se va a utilizar. Esta decisión depende de diversos factores como: la información disponible relevante al tema de estudio, el marco de muestreo factible, los objetivos de la encuesta así como los recursos económicos y humanos con los que se cuenta. Con base en estos elementos, se debe decidir sobre el tamaño de muestra; el esquema de muestreo: con igual o diferente probabilidad, número de etapas y unidades de muestreo, estratificar o no; así como elegir los estimadores adecuados con el diseño adoptado. En ocasiones las consideraciones anteriores dan lugar a más de un esquema posible, por lo que es de utilidad hacer una evaluación entre los diseños factibles.

En el caso particular de los conteos rápidos, las unidades de selección[1] son las secciones, es decir, conglomerados de votantes. No obstante, debe tomarse en cuenta que cuando el diseño de muestreo involucra selección de conglomerados de diferente tamaño, esto puede aumentar la varianza de los estimadores. En la literatura se señalan diferentes formas de incrementar la precisión del estimador, entre las que se encuentran utilizar estimadores de razón y/o muestrear con probabilidad proporcional al tamaño. En esta tesina interesa conocer si estas opciones disminuyen la varianza en conteos rápidos a nivel municipal[2]. Para ello, se compararon distintos esquemas de muestreo:

a) Estratificado, selección de conglomerados con probabilidad proporcional al tamaño y reemplazo.
b) Estratificado, selección de conglomerados con igual probabilidad y reemplazo.
c) Estratificado, selección de conglomerados con igual probabilidad y sin reemplazo.
d) Selección de conglomerados con probabilidad proporcional al tamaño y reemplazo.
e) Selección de conglomerados con igual probabilidad y reemplazo.
f) Selección de conglomerados con igual probabilidad y sin reemplazo.

[1] En el caso de un muestreo en varias etapas, las secciones corresponden a las últimas unidades de selección.

[2] Es importante mencionar que de los municipios estudiados Nezahualcóyotl, Tlalnepantla, Toluca y Valle de Chalco estaban compuestos fundamentalmente por secciones clasificados como urbanas y en términos de las preferencias electorales fueron competidos o pluripartidistas. En cambio en los municipios más pequeños: Atlacomulco, Ixtlahuaca y Jilotepec, predominó el porcentaje de secciones rurales y también una fuerza electoral

La comparación se hizo principalmente en cuanto al sesgo, la varianza y el efecto de diseño. Es importante hacer notar que las siguientes conclusiones se obtuvieron con un número fijo de secciones por municipio. No obstante, sería de gran utilidad hacer los cálculos respectivos para diferentes tamaños de muestra, quizá menores, así como también utilizar otros criterios de estratificación.

Los resultados que se obtuvieron se pueden resumir en los siguientes puntos:

- Se obtuvo una precisión mayor o igual al estratificar que cuando no se hizo.

- En términos del sesgo, el muestreo con probabilidad proporcional fue el más favorecido, aunque en todos los casos este fue pequeño (menor de 1%).

- En lo que se refiere a la varianza para los muestreos estratificados, el más preciso fue el muestreo con igual probabilidad y sin reemplazo, siendo más eficiente con respecto a los otros dos esquemas cuando la fracción de muestreo fue grande. El muestreo con diferentes probabilidades resultó más eficiente que el de igual probabilidad y reemplazo.

- Un factor que contribuyó a la ganancia del muestreo con probabilidad proporcional respecto al de igual probabilidad (ambos casos con reemplazo), fue el coeficiente de variación del tamaño de las secciones, ya que era pequeño en la mayoría de los municipios. Para los municipios en los que el coeficiente de variación no era tan pequeño se nota una mayor precisión con el ppt. Otro punto a considerar, como lo sugieren los diversos textos, es que las ganancias son moderadas cuando se estiman porcentajes.

- Respecto a los efectos de diseño, debe tenerse presente:

  Las conclusiones son análogas a las obtenidas en términos de la eficiencia, aunque los resultados no son tan evidentes. Por lo tanto, si bien es cierto que su interpretación es sencilla y son de gran utilidad práctica, sobre todo para la planeación de encuestas posteriores, se debe de tomar en cuenta que para calcular la varianza del m.a.s.s.r. se necesita el tamaño de muestra esperado bajo el muestreo complejo. De ahí que una alternativa, es usar la segunda definición del efecto de diseño, es decir, comparar el diseño complejo respecto a un muestreo aleatorio simple con reemplazo.

  Por otra parte, los efectos varían entre partidos y entre municipios, por lo que debería de tenerse cuidado al utilizarlos para calcular la varianza, o bien, el tamaño de muestra de futuros conteos rápidos

Pero ¿cuál diseño sería más apropiado utilizar? La respuesta no es única, depende de las características propias de cada dominio, y como en toda encuesta, se debe efectuar un análisis profundo que involucre diversos factores como:

- El nivel de resultados: nacional, estatal, municipal, etc. Aunque aquí sólo se analizaron encuestas municipales, sería de interés analizar mayores niveles de agregación. En principio, se esperaría que si existe mayor heterogeneidad en la población, las ganancias con una estratificación adecuada fueran mayores, ya sea en términos de precisión, costos, recursos humanos, etc.

- Las características propias del dominio de estudio. Como por ejemplo composición urbano rural, dispersión de la población, grado de escolaridad, etc.

- La fracción de muestreo. Cuando la fracción de muestreo es pequeña (lo cual puede ser el caso de un conteo rápido a nivel estatal), un muestreo sin reemplazo se ve menos favorecido por ella que cuando la fracción es considerable. Si la fracción es chica, la eficiencia de un ppt (con reemplazo) puede ser mayor o igual que la del muestreo sin reemplazo, además de las ganancias en cuanto al sesgo y al coeficiente de variación del tamaño de los conglomerados. En caso contrario, si la fracción es "grande", tal vez sería preferible un muestreo sin reemplazo e igual probabilidad. Otra opción sería un muestreo con probabilidad proporcional y sin reemplazo, que en épocas recientes ha recibido mayor impulso y desarrollo, aunque se debe estar consciente de la complicación en el cálculo de los estimadores.

- La información a priori de la que se dispone. Como por ejemplo variables para estratificar o bien estimaciones de la tendencia en las preferencias, mediante encuestas preelectorales, lo cual ayudaría a determinar la precisión deseada en las estimaciones.

- Los recursos económicos, técnicos y humanos disponibles. Quizás un diseño con costo bajo involucre un muestreo complejo, lo cual implica también tener un buen soporte de cómputo y de especialistas en la materia.

- Considerar los errores no de muestreo en que incurriría bajo cada esquema de muestreo.

Así pues, se deberá de hacer un balance entre estos elementos para obtener el diseño más adecuado.

Aunque esta decisión implica consideraciones técnicas y económicas, se debe estar consciente que también conlleva una parte de carácter ético: contribuir a que la credibilidad en las encuestas aumente o tristemente, disminuya. En la manera que se utilicen los esquemas de muestreo más propios en encuestas electorales, haciendo saber sus limitaciones y alcances, se ayudará a una mayor transparencia en los resultados electorales.

ANEXO A. VARIANZA DEL ESTIMADOR DE RAZON

A continuación se obtendrá una aproximación a la varianza[1] del estimador de razón r.

Sea

$$r = \frac{y}{x} \text{ un estimador de } R = \frac{Y}{X}$$

donde

$$Y = \sum_{i=1}^{M} Y_i$$
$$X = \sum_{i=1}^{M} X_i$$

y, x son estimadores insesgados de Y y X, respectivamente

se supone que $\frac{x - X}{X}$ es muy pequeño.

Entonces:

$$Var(r) \doteq \frac{1}{X^2}(Var(y) + R^2 Var(x) - 2RCov(y,x)).$$

Demostración:

Sean

$$\delta_y = y - E(y) = y - Y$$
$$\delta_x = x - E(x) = x - X$$

luego,

$$r - R = \frac{y}{x} - R\frac{x}{x} = \frac{\delta_y + Y - R(\delta_x + X)}{x}$$

$$= \frac{\delta_y - R\delta_x}{x} = \frac{\delta_y - \delta_x}{X}(1 + \frac{\delta_x}{X})^{-1}.$$

[1] Kish, Leslie. *Muestreo de Encuestas*. 1ª edición. Trillas, México. p. 247.

Si $\frac{\delta_x}{X}$ es pequeño, entonces

$$r - R \doteq \frac{\delta_y - R\delta_x}{X}$$

Ahora bien,

$$Var(r - R) \doteq E(r - R)^2 = \frac{E(\delta_y - R\delta_x)^2}{X^2}$$

$$= \frac{E(\delta_y^2) - 2R\delta_y\delta_x + R^2 E(\delta_x^2)}{X^2}$$

$$\doteq \frac{1}{X^2}(Var(y) + R^2 Var(x) - 2RCov(y,x)).$$

Las expresiones para r, $Var(y)$, $Var(x)$ y $Cov(y,x)$ dependen del esquema de muestreo.

A.1. Muestreo aleatorio de conglomerados sin reemplazo.

Sean

$$y = \frac{M}{m}\sum_{i=1}^{m} Y_i$$

$$x = N_m = \frac{M}{m}\sum_{i=1}^{m} N_i$$

donde:

Y_i Total de la característica de interés y en el conglomerado i

N_i Tamaño del conglomerado i

entonces, y, x son estimadores insesgados de Y y X, respectivamente, cuya varianza y covarianza están dados por

$$Var(y) = M^2 \frac{(M-m)}{M}\frac{1}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{M-1}$$

$$Var(N_m) = M^2 \frac{(M-m)}{M}\frac{1}{m}\frac{\sum_{i=1}^{M}(N_i - \frac{N}{M})^2}{M-1}$$

$$Cov(y, N_m) = M^2 \frac{(M-m)}{M} \frac{1}{m} \frac{\sum_{i=1}^{M} (Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{M-1}.$$

Prueba:

$$E(y) = \frac{M}{m} E(\sum_{i=1}^{m} Y_i)$$

$$= \frac{M}{m} \sum_{i=1}^{m} E(Y_i)$$

$$= \frac{M}{m} \sum_{i=1}^{m} \sum_{i=1}^{M} \frac{1}{M} Y_i$$

$$= \frac{M}{m} \sum_{i=1}^{m} \frac{Y}{M} = Y.$$

Análogamente,

$$E(N_m) = N$$

Veamos cuál es la varianza de y

$$Var(y) = Var(M\bar{y}_m) = M^2 Var(\bar{y}_m),$$

ya que los conglomerados se eligieron aleatoriamente y sin reemplazo, la varianza de $\bar{y}_m$ es análoga a la de un m.a.s.s.r. Entonces,

$$Var(\bar{y}_m) = \frac{(1-f)}{m} \frac{\sum_{i=1}^{M} (Y_i - \frac{Y}{M})^2}{M-1}$$

donde

$$f = m / M.$$

Por lo tanto,

$$Var(y) = M^2 \frac{(M-m)}{M} \frac{1}{m} \frac{\sum_{i=1}^{M} (Y_i - \frac{Y}{M})^2}{M-1}.$$

De manera similar

$$Var(N_m) = M^2 \frac{(M-m)}{M}\frac{1}{m}\frac{\sum_{i=1}^{M}(N_i - \frac{N}{M})^2}{M-1}$$

$$\begin{aligned}
Cov(y, N_m) &= Cov(M\bar{y}_m, M\bar{N}_m) \\
&= E((M\bar{y}_m - Y)(M\bar{N}_m - N)) \\
&= M^2 E((\bar{y}_m - \frac{Y}{M})(\bar{N}_m - \frac{N}{M})) \\
&= M^2 \frac{(1-f)}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{M-1} \\
&= M^2 \frac{(M-m)}{M}\frac{1}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{M-1}.
\end{aligned}$$

Ya que x y y son estimadores insesgados de X y Y, entonces una aproximación para la varianza del estimador de razón es:

$$Var(r) \doteq \frac{1}{N^2}(Var(y) + R^2 Var(x) - 2RCov(y,x))$$

donde

$$R = \frac{\sum_{i=1}^{M} Y_i}{\sum_{i=1}^{M} N_i} = \frac{Y}{N}$$

A.2. Muestreo aleatorio de conglomerados con reemplazo

Sean

$$y = \frac{M}{m}\sum_{i=1}^{m} Y_i$$

$$x = N_m = \frac{M}{m}\sum_{i=1}^{m} N_i$$

entonces, y, x son estimadores insesgados de Y y X, respectivamente, con varianzas

$$Var(y) = \frac{M}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{M-1}$$

$$Var(N_m) = \frac{M}{m}\frac{\sum_{i=1}^{M}(N_i - \frac{N}{M})^2}{M-1}$$

$$Cov(y, N_m) = \frac{M}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{M-1}.$$

Prueba:

$$E(y) = \frac{M}{m}E(\sum_{i=1}^{m} Y_i)$$

$$= \frac{M}{m}\sum_{i=1}^{m} E(Y_i)$$

$$= \frac{M}{m}\sum_{i=1}^{m}\sum_{i=1}^{M}\frac{1}{M}Y_i$$

$$= \frac{M}{m}\sum_{i=1}^{m}\frac{Y}{M} = Y.$$

Análogamente,

$$E(N_m) = N$$

La varianza de y es

$$Var(y) = Var(M\bar{y}_m) = M^2 Var(\bar{y}_m).$$

Ya que los conglomerados se eligieron aleatoriamente y con reemplazo, la varianza de $\bar{y}_m$ es análoga a la de un m.a.s.c.r. Por lo tanto,

$$Var(\bar{y}_m) = \frac{M-1}{M}\frac{1}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{M-1}$$

$$Var(y) = M^2 \frac{M-1}{M}\frac{1}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{M-1}$$

$$= \frac{M}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})^2}{M-1}.$$

De manera análoga

$$Var(N_m) = \frac{M}{m}\frac{\sum_{i=1}^{M}(N_i - \frac{N}{M})^2}{M-1}$$

$$Cov(y, N_m) = Cov(M\bar{y}_m, M\bar{N}_m)$$

$$= \frac{M}{m}\frac{\sum_{i=1}^{M}(Y_i - \frac{Y}{M})(N_i - \frac{N}{M})}{M-1}.$$

A.3. Muestreo con probabilidad proporcional al tamaño y con reemplazo

Sean

$$y = \frac{1}{m}\sum_{i=1}^{m}\frac{Y_i}{\pi_i}$$

$$x = N_m = \frac{1}{m}\sum_{i=1}^{m}\frac{N_i}{\pi_i}$$

entonces, y, x son estimadores insesgados de Y y X, respectivamente, cuya varianza y covarianza[2] están dados por

$$Var(y) = \frac{\sum_{i=1}^{M}\pi_i(\frac{Y_i}{\pi_i} - Y)^2}{m}$$

$$Var(N_m) = \frac{\sum_{i=1}^{M}\pi_i(\frac{N_i}{\pi_i} - N)^2}{m}$$

$$Cov(y, N_m) = \frac{\sum_{i=1}^{M}\pi_i(\frac{Y_i}{\pi_i} - Y)(\frac{N_i}{\pi_i} - N)}{m}.$$

Demostración:

Sea

t_i el número de veces que se elige el conglomerado i, $t_i = 0,1,..,m$
m=1,2,..,M.

Obsérvese que la distribución conjunta de los t_i es multinomial. Por lo tanto,

$$E(t_i) = m\pi_i$$

$$Var(t_i) = m\pi_i(1-\pi_i)$$

$$Cov(t_i, t_j) = -m\pi_i\pi_j.$$

[2] Cochran, William G. *Técnicas de Muestreo* CECSA México, 1990. p 312

Entonces,

$$E(y) = E(\frac{1}{m}\sum_{i=1}^{m}\frac{Y_i}{\pi_i})$$

$$= E(\frac{1}{m}\sum_{i=1}^{M} t_i \frac{Y_i}{\pi_i})$$

$$= \frac{1}{m}\sum_{i=1}^{M} E(t_i \frac{Y_i}{\pi_i})$$

$$= \frac{1}{m}\sum_{i=1}^{M} \frac{Y_i}{\pi_i} E(t_i)$$

$$= \frac{1}{m}\sum_{i=1}^{M} \frac{Y_i}{\pi_i} m\pi_i = Y.$$

De manera semejante,

$$E(N_m) = N$$

La varianza de y es

$$Var(y) = Var(\frac{1}{m}\sum_{i=1}^{M} t_i \frac{Y_i}{\pi_i})$$

$$= \frac{1}{m^2}\left[\sum_{i=1}^{M}\left(\frac{Y_i}{\pi_i}\right)^2 Var(t_i) + 2\sum_{i=1}^{M}\sum_{j>i}^{M}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}Cov(t_i, t_j)\right]$$

$$= \frac{1}{m^2}\left[\sum_{i=1}^{M}\left(\frac{Y_i}{\pi_i}\right)^2 m\pi_i(1-\pi_i) - 2\sum_{i=1}^{M}\sum_{j>i}^{M}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j} m\pi_i\pi_j\right]$$

$$= \frac{1}{m}\left[\sum_{i=1}^{M}\left(\frac{Y_i}{\pi_i}\right)^2 \pi_i(1-\pi_i) - 2\sum_{i=1}^{M}\sum_{j>i}^{M} Y_i Y_j\right]$$

$$= \frac{1}{m}\left[\sum_{i=1}^{M}\frac{Y_i^2}{\pi_i} - \sum_{i=1}^{M} Y_i^2 - 2\sum_{i=1}^{M}\sum_{j>i}^{M} Y_i Y_j\right]$$

$$= \frac{1}{m}\left[\sum_{i=1}^{M}\frac{Y_i^2}{\pi_i} - \sum_{i=1}^{M}\sum_{j=1}^{M} Y_i Y_j\right]$$

$$= \frac{1}{m}\left[\sum_{i=1}^{M}\frac{Y_i^2}{\pi_i} - \sum_{i=1}^{M}Y_i\sum_{j=1}^{M}Y_j\right]$$

$$= \frac{1}{m}\left[\sum_{i=1}^{M}\frac{Y_i^2}{\pi_i} - Y^2\right]$$

$$= \frac{\sum_{i=1}^{M}\pi_i\left(\frac{Y_i}{\pi_i} - Y\right)^2}{m}.$$

Análogamente,

$$Var(N_m) = \frac{\sum_{i=1}^{M}\pi_i\left(\frac{N_i}{\pi_i} - N\right)^2}{m}$$

$$Cov(y, N_m) = \frac{\sum_{i=1}^{M}\pi_i\left(\frac{Y_i}{\pi_i} - Y\right)\left(\frac{N_i}{\pi_i} - N\right)}{m}.$$

Ya que x y y son estimadores insesgados de X y Y, entonces una aproximación para la varianza del estimador de razón es:

$$Var(r) \doteq \frac{1}{N^2}(Var(y) + R^2 Var(x) - 2RCov(y,x)).$$

ANEXO B. ESTIMADORES

A continuación se darán las fórmulas para los estimadores que se utilizaron en este trabajo, de acuerdo con el esquema de muestreo correspondiente.

B.1. Muestreo estratificado de conglomerados con igual probabilidad y sin reemplazo

La notación que se utiliza es la siguiente:

$P(j)$ es la proporción de votos para el partido j en el municipio
$p(j)$ es la proporción estimada de votos para el partido j
$Y(j)_{hi}$ es el número de votos para el partido j en la sección i del estrato h
N_{hi} es el número de votos en la sección i del estrato h, $N_h = \sum_{i=1}^{M_h} N_{hi}$
m_h es el número de secciones en muestra del estrato h
M_h es el número de secciones del estrato h
H es el número de estratos
N es el número total de votos $N = \sum_{h=1}^{H}\sum_{i=1}^{M_h} N_{hi}$

En este tipo de esquema, el estimador[1] de la proporción de votos para cada partido se obtiene mediante

$$p(j) = \frac{y(j)}{N_m} = \frac{\sum_{h=1}^{H} y(j)_h}{\sum_{h=1}^{H} N_{m_h}} = \frac{\sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{m_h} Y(j)_{hi}}{\sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{m_h} N_{hi}}$$

cuya varianza[2] es aproximadamente

$$Var[p(j)] \doteq \frac{1}{N^2}(Var[y(j)] + P(j)^2 Var[N_m] - 2P(j)Cov[y(j), N_m])$$

donde[3]

$$Var[y(j)] = \sum_{h=1}^{H} Var[y(j)_h] = \sum_{h=1}^{H} \frac{M_h^2(M_h - m_h)}{m_h} \frac{\sum_{i=1}^{M_h}[Y(j)_{hi} - \frac{Y(j)_h}{M_h}]^2}{M_h(M_h - 1)}$$

---

[1] Hansen, Morris H. et al. *Sample Survey Methods and Theory Vol. I.* John Willey & Sons. New York, 1953. p. 316.

[2] Kish, Leslie. *Muestreo de Encuestas* 1ª edición. Trillas, México p. 231

[3] Hansen, Morris H. et. al. Op cit p. 317

$$Var[N_m] = \sum_{h=1}^{H} Var[N_{m_h}] = \sum_{h=1}^{H} \frac{M_h^2(M_h - m_h)}{m_h} \frac{\sum_{i=1}^{M_h}[N_{hi} - \frac{N_h}{M_h}]^2}{M_h(M_h - 1)}$$

$$Cov[\dot{y}(j), N_m] = \sum_{h=1}^{H} Cov[y(j)_h, N_{m_h}]$$

$$= \sum_{h=1}^{H} \frac{M_h^2(M_h - m_h)}{m_h} \frac{\sum_{i=1}^{M_h}[Y(j)_{hi} - \frac{Y(j)_h}{M_h}][N_{hi} - \frac{N_h}{M_h}]}{M_h(M_h - 1)}.$$

El sesgo[4] de la proporción estimada es aproximadamente

$$Sesgo[p(j)] = E[p(j)] - P(j) \doteq \frac{P(j)\big[Var[N_m] - Cov[y(j), N_m]\big]}{N^2}.$$

El efecto de diseño con respecto a un m.a.s.s.r. de votantes, *Deff*, se obtiene por

$$Deff[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{massr}]} = \frac{Var[p(j)]}{(1-f)S^2 / n^*}$$

con

$$f = \frac{n^*}{N}$$

$$S^2 = \frac{N}{N-1} P(1-P)$$

donde n^* es el número esperado de votantes en muestra bajo el diseño complejo, por lo que

$$n^* = E(\sum_{h=1}^{H}\sum_{i=1}^{m_h} N_{hi}) = \sum_{h=1}^{H} m_h \frac{N_h}{M_h}$$

y el efecto de diseño con respecto a un m.a.s.c.r. de votantes, $Deft^2$, corresponde a

$$Deft^2[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{mascr}]} = \frac{Var[p(j)]}{S^2 / n^*}.$$

[4] Kish, Leslie Op. cit. p.249.

El efecto de diseño con respecto a un m.a.s.s.r. de secciones, $Deff_c$, se obtiene por

$$Deff_c[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{massrcon}]}$$

donde:

$Var[p(j)_{massrcon}]$ es la varianza del estimador de $p(j)$ bajo un muestreo aleatorio de conglomerados (sin estratificar) con igual probabilidad y sin reemplazo.

El coeficiente de variación del estimador del número total de votos N_m es

$$C(N_m) = \frac{\sqrt{Var(N_m)}}{E(N_m)} = \frac{\sqrt{Var(N_m)}}{N}$$

B.2 Muestreo estratificado de conglomerados con igual probabilidad y con reemplazo

Se utiliza la misma notación que para el muestreo sin reemplazo. Además, el estimador de la proporción de votos por partido es igual al caso anterior, y el de la varianza es similar, excepto por un factor

$$p(j) = \frac{y(j)}{N_m} = \frac{\sum_{h=1}^{H} y(j)_h}{\sum_{h=1}^{H} N_{m_h}} = \frac{\sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{m_h} Y(j)_{hi}}{\sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{m_h} N_{hi}}$$

su varianza es

$$Var[p(j)] \doteq \frac{1}{N^2}(Var[y(j)] + P(j)^2 Var[N_m] - 2P(j)Cov[y(j), N_m])$$

donde

$$Var[y(j)] = \sum_{h=1}^{H} Var[y(j)_h] = \sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{M_h} [Y(j)_{hi} - \frac{Y(j)_h}{M_h}]^2$$

$$Var[N_m] = \sum_{h=1}^{H} Var[N_{m_h}] = \sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{M_h} [N_{hi} - \frac{N_h}{M_h}]^2$$

$$Cov[y(j), N_m] = \sum_{h=1}^{H} Cov[y(j)_h, N_{m_h}] = \sum_{h=1}^{H} \frac{M_h}{m_h} \sum_{i=1}^{M_h} [Y(j)_{hi} - \frac{Y(j)_h}{M_h}][N_{hi} - \frac{N_h}{M_h}].$$

El sesgo de la proporción estimada se calcula por

$$Sesgo[p(j)] = E[p(j)] - P(j) \doteq \frac{P(j)\big[Var[N_m] - Cov[y(j), N_m]\big]}{N^2}.$$

El $Deff$ se define como

$$Deff[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{massr}]} = \frac{Var[p(j)]}{(1-f)S^2 / n*}$$

donde

$$f = \frac{n*}{N}$$

$$S^2 = \frac{N}{N-1} P(1-P)$$

n^* es el número esperado de votantes en muestra bajo el diseño señalado, entonces

$$n* = E(\sum_{h=1}^{H} \sum_{i=1}^{m_h} N_{hi}) = \sum_{h=1}^{H} m_h \frac{N_h}{M_h}$$

el $Deft^2$ corresponde a

$$Deft^2[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{mascr}]} = \frac{Var[p(j)]}{S^2 / n*}.$$

El $Deff_c$ se obtiene mediante

$$Deff_c[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{massrcon}]}$$

donde:

$Var[p(j)_{masrcon}]$ es la varianza del estimador de $p(j)$ bajo un muestreo aleatorio de conglomerados con igual probabilidad y sin reemplazo.

El coeficiente de variación N_m es

$$C(N_m) = \frac{\sqrt{Var(N_m)}}{E(N_m)} = \frac{\sqrt{Var(N_m)}}{N}$$

B.3. Muestreo estratificado de conglomerados con probabilidad proporcional al tamaño y con reemplazo

Notación:

$P(j)$ es la proporción de votos para el partido j en el municipio

$p(j)_{ppt}$ es la proporción estimada de votos para el partido j

$Y(j)_{hi}$ es el número de votos para el partido j en la sección i del estrato h

N_{hi} es el número de votos en la sección i del estrato h, $N_h = \sum_{i=1}^{M_h} N_{hi}$

π_{hi} es la probabilidad de seleccionar la sección i dentro del estrato h $\pi_{hi} = \frac{lista_{hi}}{lista_h}$

m_h es el número de secciones en muestra del estrato h

M_h es el número de secciones del estrato h

H es el número de estratos

N es el número total de votos $N = \sum_{h=1}^{H} \sum_{i=1}^{M_h} N_{hi}$

El estimador[5] de la proporción de votos por partido se puede expresar como

$$p(j)_{ppt} = \frac{y(j)_{ppt}}{N_{m,ppt}} = \frac{\sum_{h=1}^{H} y(j)_{h,ppt}}{\sum_{h=1}^{H} N_{m_h,ppt}} = \frac{\sum_{h=1}^{H} \frac{1}{m_h} \sum_{i=1}^{m_h} Y(j)_{hi} / \pi_{hi}}{\sum_{h=1}^{H} \frac{1}{m_h} \sum_{i=1}^{m_h} N_{hi} / \pi_{hi}}$$

[5] Hansen, Morris H. et. al. Op cit p 417.

cuya varianza es aproximadamente

$$Var[p(j)_{ppt}] \doteq \frac{1}{N^2}(Var[y(j)_{ppt}] + P(j)^2 Var[N_{m,ppt}] - 2P(j)Cov[y(j)_{ppt}, N_{m,ppt}])$$

donde[6]

$$Var[y(j)_{ppt}] = \sum_{h=1}^{H} Var[y(j)_{h,ppt}] = \sum_{h=1}^{H} \frac{\sum_{i=1}^{M_h} \pi_{hi}[\frac{Y(j)_{hi}}{\pi_{hi}} - Y(j)_h]^2}{m_h}$$

$$Var[N_{m,ppt}] = \sum_{h=1}^{H} Var[N_{m_h,ppt}] = \sum_{h=1}^{H} \frac{\sum_{i=1}^{M_h} \pi_{hi}(\frac{N_{hi}}{\pi_{hi}} - N_h)^2}{m_h}$$

$$Cov[y(j)_{ppt}, N_{m,ppt}] = \sum_{h=1}^{H} Cov[y(j)_{h,ppt}, N_{m_h,ppt}] = \sum_{h=1}^{H} \frac{\sum_{i=1}^{M_h} \pi_{hi}[\frac{Y(j)_{hi}}{\pi_{hi}} - Y(j)_h][\frac{N_{hi}}{\pi_{hi}} - N_h]}{m_h}$$

Una aproximación para el sesgo de la proporción estimada es

$$Sesgo[p(j)_{ppt}] = E[p(j)_{ppt}] - P(j) \doteq \frac{P(j)\left[Var[N_{m,ppt}] - Cov[y(j), N_{m,ppt}]\right]}{N^2}.$$

El efecto de diseño *Deff* se obtiene mediante

$$Deff\,[p(j)_{ppt}] = \frac{Var[p(j)_{ppt}]}{Var[p(j)_{mas\,sr}]} = \frac{Var[p(j)_{ppt}]}{(1-f)S^2/n^*}$$

con

$$f = \frac{n^*}{N}$$

$$S^2 = \frac{N}{N-1}P(1-P)$$

[6] Hansen, Morris H. et al Op cit. p 417.

donde n^* es el número esperado de votantes en muestra bajo el diseño complejo, por lo que

$$n^* = E(\sum_{h=1}^{H}\sum_{i=1}^{m_h} N_{hi}) = \sum_{h=1}^{H} m_h \sum_{i=1}^{M_h} \pi_{hi} N_{hi}$$

y el efecto de diseño $Deft^2$ corresponde a

$$Deft^2[p(j)_{ppt}] = \frac{Var[p(j)_{ppt}]}{Var[p(j)_{mascr}]} = \frac{Var[p(j)_{ppt}]}{S^2 / n^*}$$

El efecto de diseño, $Deff_c$, se obtiene mediante

$$Deff_c[p(j)] = \frac{Var[p(j)]}{Var[p(j)_{massrcon}]}$$

donde:

$Var[p(j)_{massrcon}]$ es la varianza del estimador de $p(j)$ bajo un muestreo aleatorio de conglomerados con igual probabilidad y sin reemplazo.

El coeficiente de variación del estimador del número total de votos $N_{m,ppt}$ es el siguiente

$$C(N_{m,ppt}) = \frac{\sqrt{Var(N_{m,ppt})}}{E(N_{m,ppt})} = \frac{\sqrt{Var(N_{m,ppt})}}{N}.$$

En general, la eficiencia relativa de un estimador $\hat{\theta}_1$ con respecto a otro estimador $\hat{\theta}_2$, se define mediante:

$$Eficiencia = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}.$$

Así, la eficiencia relativa del muestreo ppt en relación al muestreo de igual probabilidad y con reemplazo se obtiene por

$$\frac{Var[p(j)_{cr}]}{Var[p(j)_{ppt}]}.$$

Análogamente, la eficiencia relativa del muestreo ppt con respecto al muestreo de igual probabilidad y sin reemplazo se obtiene mediante

$$\frac{Var[p(j)_{sr}]}{Var[p(j)_{ppt}]}.$$

Debe hacerse notar que cuando se hable de porcentajes en lugar de proporciones, las fórmulas referentes a $P(j)$, $p(j)$ y $Sesgo[p(j)]$, bajo el respectivo muestreo, deben de multiplicarse por 100. Las fórmulas para las varianzas ,$Var[p(j)], Var[p(j)_{mas}]$, deberán multiplicarse por 100^2.

Cuando no se estratifica H=1.

BIBLIOGRAFIA

COCHRAN, W.G. *Técnicas de Muestreo*. Tr. Andrés Sestier Bouclier. C.E.C.S.A. México, 1980. 513 p.

CONSULTA. *Reporte Metodológico: Elecciones en el Estado de México*. México, 1996.

FOREMAN, E.K. *Survey Sampling Principles*. MARCEL DEKKER. New York, 1991. 497 p.

HANSEN, M.H., HURWITZ, W.N. y MADOW, W.G. *Sample Survey Methods and Theory Vols. I y II*. John Willey & Sons. New York, 1953. 638 y 332 p.

KISH, L. *Muestreo de Encuestas*. Tr. Ricardo Vinós Cruz López. Trillas. México, 1972. 739 p.

KISH, L. *Defs: why, when and how? a review*. ASA Proceedings of Survey Research Methods Section, 1989. 209-211 p.

LEYVA, E. N. *El Muestreo para Proporciones. Un Caso Práctico en Encuestas para Medición de Preferencias Electorales*. Tesis. Escuela Nacional de Estudios Profesionales Acatlán, UNAM. México, 1995. 113 p.

SUKHATME, P.V., SUKHATME, B.V. Y ASOK, C. *Sampling Theory of Surveys with Applications*, 3ª edición. Iowa State University Press. Iowa, 1984. 526 p.

WOLTER, K.M. *Introduction to Variance Estimation*. Springer-Verlang. New York, 1985. 427 p.