

03061



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

1

UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONAL Y POSGRADO
DEL COLEGIO DE CIENCIAS Y HUMANIDADES

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y
SISTEMAS

**Evaluación de errores de muestreo en conteos
rápidos para elecciones presidenciales**

T E S I S

QUE PARA OBTENER EL GRADO DE
MAESTRO EN ESTADÍSTICA E
INVESTIGACIÓN DE OPERACIONES

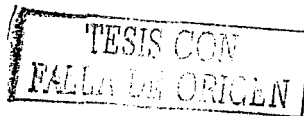
P R E S E N T A:

JOSÉ LUIS CASTREJÓN CABALLERO

castrejon@correo.unam.mx

DIRECTORA DE TESIS: DRA. GUILLERMINA ESLAVA GÓMEZ

MÉXICO, D.F.



2003

A



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicatorias

A Tere:

Este trabajo es de los dos, gracias por estos diez años, por tu ayuda y por confiar en mí, eres una gran mujer.

A mi hermana Mary:

Quien ha soportado mi mal humor en los últimos años. Gracias por todo y espero te vaya mejor en esta nueva etapa de tu vida.

A mis hermanos Juan, Carlos, Ramón, Sergio, Moisés y Urbano:

Se que esto es también muy importante para ustedes

A mis alumnos pasados, presentes y futuros:

Son la motivación de mi superación.

A Norita, Tere, Héctor, Ricardo y Mario:

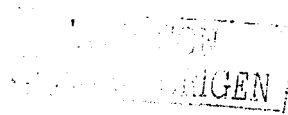
Gracias por dejarme compartir con ustedes un poco de tiempo y de sus discusiones.

A mi Mamá:

Sin tí no podría hacer nada. Te debo todo.

A mis sobrinos Juanito, Omar, Janet, Miriam, Fabián, Nayelli, Verito, Gaby, Daniel, Alondra, Moy y Selene:

Son la vida misma.



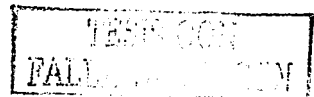
Agradecimientos

Agradezco en primer lugar al pueblo de México que vía sus impuestos sostiene ese gran proyecto llamado educación pública, muchos no podríamos llegar a esta etapa si no existiera. Para realizar los estudios de maestría tuve la fortuna de contar con el apoyo del CONACYT que me otorgo la beca con registro 121431. La UNAM no solo me permitió obtener educación gratuita y de alta calidad, también me otorgo un complemento de beca a través de la Dirección General de Estudios de Posgrado. Reconozco el apoyo brindado por la Subcomisión de Capacitación Mixta de los trabajadores ATM del INAH para la realización de los estudios de maestría.

Esta tesis no hubiera sido posible sin las ideas, comentarios y atinada dirección de la Dra. Guillermina Eslava Gómez con quien quedo en deuda permanente, le agradezco su paciencia y apoyo que me ha brindado en toda mi formación académica.

Agradezco a la Mtra. Patricia Romero, a la Dra. Rebeca Aguirre, al Dr. Ignacio Méndez y al Dr. Carlos Díaz los comentarios y sugerencias de este trabajo y por aceptar ser miembros del jurado del examen de grado.

Reconozco la calidad académica de mis profesores de la maestría: Dra. Guillermina Eslava, Dra. Silvia Ruiz-Velasco, Mtra. Leticia Gracia, Mtra. Patricia Romero, Dr. Raúl Rueda, Dr. Federico O' Reilly, Mtro. Salvador Zamora, Dr. Mogens Bladt y Dr. Eduardo Gutiérrez.



Agradezco de manera particular al Mtro. Salvador Zamora, quien además de ser un excelente maestro es un gran amigo y a la Mtra. Patricia Romero por el apoyo y comentarios brindados al inicio de este trabajo.

A mis colegas estudiantes de la maestría Mónica, Paty, Yazmín, Oswaldo, Jaime, Chucho y Karim les agradezco la convivencia y apoyo en los dos años que coincidimos.

Reconozco el apoyo de las compañeras Juanita y Cecy de la biblioteca del IIMAS por su apoyo y charla. Asimismo doy gracias a la otra Juanita por el apoyo en fotocopias y su animo permanente. A Alfredo por la ayuda para todos los trámites que he realizado en mis estudios en el IIMAS. A Eli del Departamento de Probabilidad y Estadística del IIMAS, muchas gracias por todo.

Agradezco a mis colegas y profesores de la ENAH, Dr. Sergio López y Dr. Lauro González, así como a varios estudiantes de la licenciatura de Antropología Física de la ENAH sus palabras de aliento para realizar esta tesis.

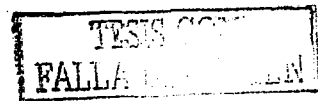


Resumen

Se evalúan los errores de muestreo esperados en conteos rápidos para las elecciones presidenciales del 2000 en México. Estos errores se evalúan para diversos tamaños de muestra, bajo diversos diseños muestrales y usando el estimador de razón.

Se trabaja con los resultados de las votaciones para elección a presidente del 2 de julio del 2000, desagregados a nivel sección y para la totalidad del país. Se consideran tamaños de muestra que varían desde 300 hasta 2000 secciones, esto para llevar al extremo el comportamiento de los errores; aun cuando para fines prácticos y operativos, los tamaños de muestra recomendables varían desde 600 hasta 1500 secciones. Los diversos diseños muestrales considerados se pueden clasificar en tres grupos: de una etapa, estratificados de una etapa, y bietápicos. La evaluación de estos errores, ofrece una guía para la selección de diseño y tamaño de muestra para conteos rápidos a nivel nacional, ofrece una gama de posibilidades y permite apreciar que no existe un diseño único y mejor, sino que la logística en campo, los tiempos, y los recursos disponibles ayudan a seleccionar un tamaño de muestra y un diseño muestral adecuado.

Se hacen también observaciones sobre la variabilidad de las respuestas de los votantes, tanto al interior de los distritos electorales como entre distritos electorales, para cada partido. Este trabajo permite hacer una evaluación objetiva, de observaciones cualitativas que se han expresado sobre la bondad y eficacia de los múltiples conteos rápidos llevados a la práctica ese 2 de julio.

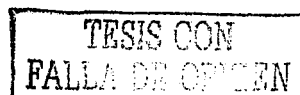


Índice

Introducción	1
Capítulo 1. Diseños de Muestreo	4
1.1 Introducción	4
1.2 Muestreo aleatorio simple sin reemplazo	4
1.3 Muestreo con probabilidad proporcional al tamaño	7
a) Con reemplazo	7
b) Sin reemplazo	8
1.4 Muestreo estratificado	8
1.4.1 Estimación separada	9
1.4.2 Estimación combinada	10
1.5 Muestreo bietápico	10
1.5.1 Muestreo aleatorio simple en ambas etapas	11
1.5.2 Muestreo PPT en la primera etapa y muestreo aleatorio simple en la segunda etapa	12
a) Unidades primarias de muestreo seleccionadas con reemplazo	12
b) Unidades primarias de muestreo seleccionadas sin reemplazo	12
1.6 Sesgo y Coeficiente de Variación	13
Capítulo 2. Análisis Exploratorio del Listado Nominal de las Secciones Electorales	14
2.1 Composición del padrón electoral	14
2.2 Resultado general de la elección	20
2.3 Resultados por circunscripción	21
2.4 Resultados según tipo de sección	29
2.5 Descripción multivariada por estado	30
Capítulo 3: Resultados	34
3.1 Notación	34
3.1.1 Muestreo aleatorio simple	34
3.1.2 Muestreo con probabilidad proporcional al tamaño	35
a) Con reemplazo	35
b) Sin reemplazo	35
3.1.4 Muestreo estratificado. Estimador simple	35

TRABAJO CON
FOLIO DE ORIGEN

3.1.5 Muestreo estratificado. Estimador combinado	36
3.1.6 Muestreo bietápico	36
a) Muestreo aleatorio simple en ambas etapas	37
b) Muestreo con probabilidad proporcional al tamaño en la primera etapa con reemplazo y aleatorio simple en la segunda etapa	38
c) Muestreo con probabilidad proporcional al tamaño en la primera etapa sin reemplazo y aleatorio simple en la segunda etapa	38
3.1.8 Sesgo y coeficiente de variación en el muestreo aleatorio simple	38
3.2 Error cuadrático medio	39
3.3 Muestreo bietápico	54
3.4 Error de estimación centrado en el parámetro para los diferentes diseños muestrales	66
3.5 Sesgo en el muestreo aleatorio simple	67
Capítulo 4. Conclusiones	76
Anexo: Código de Programas en Fortran	80
A.1 Muestreo Aleatorio Simple, Muestreo Proporcional al Tamaño sin Reemplazo y con Reemplazo	80
A.2 Muestreo Estratificado, Estimador Separado y Combinado. Caso de Circunscripciones	83
A.3 Muestreo Bietápico, Distritos Electorales como UPM y Secciones como USM. Muestreo aleatorio simple en ambas etapas.	86
A.4 Muestreo Bietápico, Distritos Electorales como UPM y Secciones como USM. Muestreo PPT en la primera etapa con reemplazo y MAS en la segunda.	89
A.5 Muestreo Bietápico, Distritos Electorales como UPM y Secciones como USM. Muestreo PPT en la primera etapa sin reemplazo y MAS en la segunda.	92
A.6 Sesgo y Coeficiente de variación en el Muestreo Aleatorio Simple	94
Referencias	97



Índice de Cuadros y Figuras

Cuadro 2.1 Distribución de secciones, lista nominal y distritos, por circunscripción y estado.	16
Figura 2.1 Porcentaje de electores registrados por estado según lista nominal.	17
Cuadro 2.2 Distribución de las secciones electorales según tipo a nivel nacional.	18
Cuadro 2.3 Distribución de lista nominal por tipo en cada entidad.	19
Figura 2.2 Tipo de sección electoral: porcentajes en cada entidad.	20
Figura 2.3 Resultados de las elecciones presidenciales 2000.	21
Cuadro 2.4 Resultados de las elecciones: estados de la primera circunscripción.	22
Figura 2.4 Porcentaje de votación para cada partido en los estados de la primera circunscripción.	23
Cuadro 2.5 Resultados de las elecciones: estados de la segunda circunscripción.	24
Figura 2.5 Porcentaje de votación para cada partido en los estados de la segunda circunscripción.	25
Cuadro 2.6 Resultados de las elecciones: estados de la tercera circunscripción.	25
Figura 2.6 Porcentajes de votación para cada partido en los estados de la tercera circunscripción.	26
Cuadro 2.7 Resultados de las elecciones: estados de la cuarta circunscripción.	27
Figura 2.7 Porcentaje de votación para cada partido en los estados de la cuarta circunscripción.	27
Cuadro 2.8 Resultados de las elecciones: estados de la quinta circunscripción.	28

Figura 2.8 Porcentaje de votación para cada partido en los estados de la cuarta circunscripción.	28
Figura 2.9 Entidades según voto mayoritario para los diferentes partidos políticos.	29
Figura 2.10 Distribución de votos según tipo de sección electoral.	30
Cuadro 2.9 Coeficientes de las componentes principales y las proporciones de los partidos por estado.	31
Figura 2.11 Gráfica de los dos primeros componentes principales.	32
Cuadro 2.10 Claves por estado, usados en dendograma.	32
Figura 2.12 Dendograma que muestra la agrupación de los estados según la votación obtenida, utilizando análisis de conglomerados jerárquica.	33
Cuadro 3.1 Notación para los diseños de muestreo utilizados.	40
Cuadro 3.2 Error Cuadrático Medio para diferentes tamaños de muestra, considerando diferentes diseños de muestreo.	42
Figura 3.1 Error de estimación para varios tamaños de muestra, bajo diferentes diseños de muestreo. Alianza por el Cambio.	47
Figura 3.2 Error de estimación para varios tamaños de muestra, bajo diferentes diseños de muestreo. PRI.	48
Figura 3.3 Error de estimación para varios tamaños de muestra, bajo diferentes diseños de muestreo. Alianza por México	49
Figura 3.4 Error de estimación para varios tamaños de muestra, bajo diferentes diseños de muestreo. Otros partidos	50
Figura 3.5 Error de estimación para diseños bietápicos varios tamaños de UPM, con 4 USM. Alianza por el Cambio.	51
Figura 3.6 Error de estimación para diseños bietápicos varios tamaños de UPM, con 4 USM. PRI.	52
Figura 3.7 Error de estimación para diseños bietápicos varios tamaños de UPM, con 4 USM. Alianza por México.	53

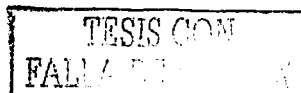


Figura 3.8 Error de estimación, en el muestreo bietápico (BIEMAS), con Distritos como UPM y Secciones como USM, para los diferentes partidos.	56
Figura 3.9 Error de estimación correspondiente a Unidades Primarias de Muestreo (Distritos), primer término de la expresión 1.8, en el muestreo bietápico (BIEMAS), para los diferentes partidos.	57
Figura 3.10 Error de estimación correspondiente a Unidades Secundarias de Muestreo (Secciones), segundo término de la expresión 1.8, en el muestreo bietápico (BIEMAS), para los diferentes partidos.	58
Figura 3.11 Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM(Distritos) y de USM(Secciones). Alianza por el Cambio	59
Figura 3.12 Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM (Distritos) y de USM (Secciones). PRI	60
Figura 3.13 Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM (Distritos) y de USM (Secciones). Alianza por México.	61
Figura 3.14 Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM (Distritos) y de USM (Secciones). Otros partidos.	62
Figura 3.15 Error cuadrático en el muestreo bietápico (BIEMAS), para unidades primarias y secundarias de muestreo (2 y 3 USM en cada UPM). Alianza por el Cambio.	63
Figura 3.16 Error cuadrático en el muestreo bietápico para unidades primarias y secundarias de muestreo (2 y 3 USM en cada UPM).PRI	64
Figura 3.17 Error cuadrático en el muestreo bietápico (BIEMAS) para unidades primarias y secundarias de muestreo (2 y 3 USM en cada UPM). Alianza por México.	65
Figura 3.18 Error de estimación en el muestreo bietápico (BIEMAS), para unidades primarias y secundarias de muestreo (2 a 5 USM en cada UPM).	69

TESTS CON
FOLIO DE ORIGEN

Figura 3.19. Error de estimación centrado en el parámetro de interés de cada partido, bajo diferentes diseños de muestreo.	70
Figura 3.20 Error de estimación centrado en el parámetro de interés, bajo diferentes diseños de muestreo. Alianza por el cambio.	71
Figura 3.21 Error de estimación centrado en el parámetro de interés, bajo diferentes diseños de muestreo. PRI	72
Figura 3.22 Error de estimación centrado en el parámetro de interés, bajo diferentes diseños de muestreo. Alianza por México.	73
Figura 3.23 Error de estimación centrado en el parámetro de interés de cada partido, Muestreo Bietápico (BIEMAS), diferentes tamaños de Unidades Primarias (Distritos) y Secundarias (Secciones) de Muestreo.	74
Figura 3.24 Sesgo relativo y coeficiente de variación en el muestreo aleatorio simple. Alianza por el Cambio.	75

TESIS CON
FALLA DE ORIGEN

Introducción

La necesidad de realizar encuestas en muchas de las actividades que se realizan en estos tiempos es manifiesta. En el área de la mercadotecnia es uno de los instrumentos más eficaces para medir el probable éxito de un producto o servicio. En aspectos políticos la encuesta es usada para diferentes objetivos: medir la popularidad de un político, medir la aceptación de una obra pública, determinar la candidatura para ocupar un cargo, definir cual partido o candidato lleva la preferencia en un momento previo a una elección o el mismo día de la elección o medir de una manera rápida la tendencia de los resultados unas horas después de concluir el proceso de votación el día señalado. Este último aspecto es el tema que se trata en este trabajo.

Las elecciones presidenciales del 6 de julio de 1988 fueron quizás las más cuestionadas de los últimos tiempos en México. La organización de los procesos electorales estaba a cargo del gobierno a través de la Secretaria de Gobernación. La noche de la elección las autoridades responsables argumentaron una caída del sistema de cómputo en el programa de resultados preliminares, por lo cual la población mexicana se quedó sin conocer la tendencia de los resultados esa misma noche como ya ocurría en muchos países del mundo. Fue hasta dos días después que empezaron a conocerse los resultados, lo cual despertó sospechas de fraude por parte de los partidos opositores.

Este hecho sirvió en lo político para realizar reformas a las leyes electorales de manera que los órganos encargados de realizar las elecciones fueran independientes de las autoridades gubernamentales. Estos acontecimientos también repercutieron en los aspectos técnicos relacionados con las elecciones, en cuanto a la elaboración de encuestas por parte de empresas encargadas de estudios de opinión pública. En torno a una elección existen cuatro grandes tipos de encuestas: las encuestas pre-electorales que se realizan en muestras de la población en diferentes momentos previos al día de la elección; las encuestas de salida o *exit poll* que se realizan un momento después de que una persona ha

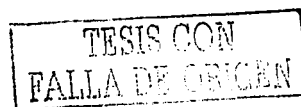


depositado su voto; los conteos rápidos que consisten en realizar un muestreo de casillas electorales (en realidad de secciones electorales) y considerar los resultados totales obtenidos en ella; y las encuestas postelectorales que se aplican unos días después de la elección. En las elecciones para Presidente de la República realizadas en 1994 y 2000 se aplicaron los diferentes tipos de encuestas por empresas e instituciones interesadas en conocer las tendencias de las preferencias electorales tiempo antes de la elección así como para conocer con prontitud las tendencias de los resultados el día de la elección.

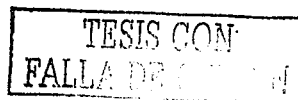
El objetivo de este trabajo es evaluar el desempeño del estimador de razón para diferentes diseños muestrales para la encuesta denominada conteo rápido. Se estudiará la eficiencia de los estimadores de razón para diferentes diseños muestrales: aleatorio simple sin reemplazo, con probabilidad proporcional al tamaño con y sin reemplazo, estratificado y bietápico.

Los datos que se utilizan para el análisis corresponden al listado nominal por cada sección electoral y los resultados oficiales de las elecciones presidenciales efectuadas el 2 de julio del 2000, proporcionados por el Instituto Federal Electoral (IFE). Una sección electoral puede estar compuesta por una casilla básica, una o más contiguas y una o más especiales. Las organizaciones participantes en la elección fueron:(1) la coalición denominada Alianza por el Cambio que estuvo integrada por el Partido Acción Nacional (PAN) y el Partido Verde Ecologista de México (PVEM); (2) el Partido Revolucionario Institucional (PRI); (3) la coalición denominada Alianza por México que se formó con el Partido de la Revolución Democrática (PRD), el Partido del Trabajo (PT), el Partido Alianza Social (PAS), el partido Convergencia por la Democracia y el Partido de la Sociedad Nacionalista (PSN); (4) el Partido del Centro Democrático; (5) el Partido Auténtico de la Revolución Mexicana (PARM) y (6) Democracia Social Partido Político Nacional (DSPPN).

La tesis esta dividida en cuatro capítulos y un anexo. En el primer capítulo se presentan las expresiones matemáticas para calcular los diferentes estimadores de razón a utilizar según el diseño de muestreo, así como la expresión del error



cuadrático medio. En el capítulo dos se presenta un análisis descriptivo de la composición del padrón electoral así como de los resultados obtenidos por las diferentes organizaciones políticas participantes. En el tercer capítulo se evalúan los resultados del error cuadrático medio de los estimadores de razón para los diseños muestrales considerados. El capítulo cuarto esta dedicado a las conclusiones y reflexiones finales del trabajo. Para los resultados que se presentan en el capítulo tercero se utilizaron programas elaborados en Fortran 90. El código de estos programas se incluye en el anexo.



Capítulo 1: Diseños de Muestreo

1.1 Introducción

Los métodos tradicionales de muestreo como el Muestreo Aleatorio Simple, el Muestreo Estratificado, el Muestreo por Conglomerados, el Muestreo Bietápico o alguna combinación de ellos, son utilizados para estimar medias, proporciones o totales de una variable en alguna población. En ocasiones se requiere estimar alguna combinación no lineal de dos variables considerando una de ellas como auxiliar. Este es el caso del estimador de razón el cual es muy útil en la práctica. A continuación se presenta el planteamiento para este estimador.

Sean (X_i, Y_i) los diferentes pares de valores relativos a la variable en estudio X y la variable auxiliar respectivamente. Se denomina razón poblacional R de X y Y al cociente entre los totales poblacionales de X y Y . Es decir:

$$R = \frac{\sum_{i=1}^N X_i}{\sum_{i=1}^N Y_i}$$

Donde N = tamaño de la población.

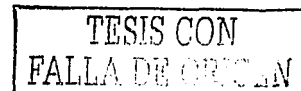
1.2 Muestreo aleatorio simple sin reemplazo

Si el diseño de muestreo realizado es el aleatorio simple sin reemplazo, el estimador de R esta dado por la expresión:

$$\hat{R} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} = \frac{\bar{x}}{\bar{y}}$$

$$\text{donde } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{y} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Con n = tamaño de muestra y (x_i, y_i) son los elementos en muestra de las dos variables.



Considerando que $E(\hat{R}) = E\left(\frac{\sum_{i=1}^{n_i} X_i}{\sum_{i=1}^{n_i} Y_i}\right) \neq \frac{E\left(\sum_{i=1}^{n_i} X_i\right)}{E\left(\sum_{i=1}^{n_i} Y_i\right)} = \frac{\sum_{i=1}^{N_i} X_i}{\sum_{i=1}^{N_i} Y_i} = R$ se puede

concluir que el estimador \hat{R} no es un estimador insesgado de R . El sesgo de este estimador se puede aproximar de acuerdo al desarrollo presentado por Raj (1984, pp. 101-103):

$$\hat{R} - R = \frac{\bar{x}}{\bar{y}} - R = \frac{\bar{x} - R\bar{y}}{\bar{y}} \cdot \frac{\bar{y}}{\bar{y}} = \frac{\bar{x} - R\bar{y}}{\bar{y}} \cdot \frac{\bar{y}}{\bar{y} + \bar{y} - \bar{y}} = \frac{\bar{x} - R\bar{y}}{\bar{y}} \cdot \frac{1}{1 + \frac{\bar{y} - \bar{Y}}{\bar{Y}}}$$

Sabiendo que si $|r| < 1 \Rightarrow \sum_{n=0}^{\infty} (-r)^n = \frac{1}{1 - (-r)} = \frac{1}{1 + r}$, se tiene: que $\left|\frac{\bar{y} - \bar{Y}}{\bar{Y}}\right| < 1$,

por lo que se tiene: $\hat{R} - R = \frac{\bar{x} - R\bar{y}}{\bar{y}} \cdot \left[1 - \frac{\bar{y} - \bar{Y}}{\bar{Y}} + \frac{(\bar{y} - \bar{Y})^2}{\bar{Y}^2} - \dots\right]$ por lo tanto como

$Sesgo(\hat{R}) = E(\hat{R}) - R = E(\hat{R} - R)$, y cortando hasta el término de primer grado la serie anterior, una aproximación para el sesgo esta dada por:

$$Sesgo(\hat{R}) \approx E\left(\frac{\bar{x} - R\bar{y}}{\bar{y}} \cdot \left[1 - \frac{\bar{y} - \bar{Y}}{\bar{Y}}\right]\right) = E\left(\frac{\bar{x} - R\bar{y}}{\bar{y}} - \frac{(\bar{x} - R\bar{y})(\bar{y} - \bar{Y})}{\bar{y}\bar{Y}}\right) = E\left(\frac{\bar{x} - R\bar{y}}{\bar{y}}\right) - E\left(\frac{(\bar{x} - R\bar{y})(\bar{y} - \bar{Y})}{\bar{y}\bar{Y}}\right)$$

$$= \frac{E(\bar{x} - R\bar{y})}{\bar{y}} - \frac{E[(\bar{x} - R\bar{y})(\bar{y} - \bar{Y})]}{\bar{y}\bar{Y}} = \frac{E[(\bar{x} - R\bar{y})(\bar{y} - \bar{Y})]}{\bar{y}\bar{Y}} = \frac{RVar(\bar{y}) - Cov(\bar{x}, \bar{y})}{\bar{y}\bar{Y}}$$

Se usaron las siguientes igualdades:

$$E(\bar{x} - R\bar{y}) = E(\bar{x}) - R \cdot E(\bar{y}) = \bar{X} - R\bar{Y} = \bar{X} - \frac{\bar{X}}{\bar{Y}}\bar{Y} = \bar{X} - \bar{X} = 0$$

$$E[(\bar{x} - R\bar{y})(\bar{y} - \bar{Y})] = E(\bar{x}\bar{y}) - RE(\bar{y}^2) + RYE(\bar{y}) = E(\bar{x}\bar{y}) - \bar{X}\bar{Y} - R(E(\bar{y}^2) - [E(\bar{y})]^2)$$

$$= cov(\bar{x}, \bar{y}) - RVar(\bar{y})$$

TESIS CON
FALLA DE CALIFICACION

Como en el muestreo aleatorio simple la varianza del estimador \bar{y} es $(1-f)\frac{S_Y^2}{n}$, donde $f = \frac{n}{N}$ es la fracción de muestreo, la expresión aproximada para el sesgo es:

$$\text{Sesgo}(\hat{R}) = \frac{\left(1 - \frac{n}{N}\right)}{n\bar{Y}^2} \left(RS_Y^2 - S_{XY}\right).$$

Debido a que $S_Y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$ y $S_{XY} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})}{N-1}$

se tiene finalmente:

$$\text{Sesgo}(\hat{R}) = \frac{\left(1 - \frac{n}{N}\right)}{n(N-1)\bar{Y}^2} \left(R \sum_{i=1}^N Y_i^2 - \sum_{i=1}^N X_i Y_i\right)$$

La varianza aproximada o Error Cuadrático Medio del estimador, tomando en cuenta sólo el primer término en el desarrollo de la serie para la expresión $\hat{R} - R$, es:

$$ECM(\hat{R}) = E(\hat{R} - R)^2 = V(\hat{R} - R) = \text{Var}\left(\frac{\bar{x} - R\bar{y}}{\bar{Y}}\right) = \frac{\text{Var}(\bar{x} - R\bar{y})}{\bar{Y}^2}$$

Por lo tanto

$$ECM(\hat{R}) = \frac{\text{Var}(\bar{x}) + R^2 \text{var}(\bar{y}) - 2RCov(\bar{x}, \bar{y})}{\bar{Y}^2} \quad (1.1)$$

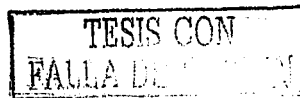
Sustituyendo las expresiones para el muestreo aleatorio simple queda:

$$ECM(\hat{R}) = \frac{(1-f)\frac{S_x^2}{n} + R^2(1-f)\frac{S_y^2}{n} - 2R(1-f)\frac{S_{XY}}{n}}{\bar{Y}^2} = \frac{1-f}{\bar{Y}^2 n(N-1)} \left(\sum_{i=1}^N X_i^2 + R^2 \sum_{i=1}^N Y_i^2 - 2R \sum_{i=1}^N X_i Y_i \right)$$

La cual se reduce a la expresión:

$$ECM(\hat{R}) = \left(1 - \frac{n}{N}\right) \frac{N^2}{n} \left(\frac{1}{N-1}\right) \frac{1}{\bar{Y}^2} \sum_{i=1}^N (X_i - RY_i)^2 \quad (1.2)$$

(Cochran, 1993, pág. 199, ec. 6.4)



1.3 Muestreo con probabilidad proporcional al tamaño

Si el muestreo se realiza con Probabilidad Proporcional al Tamaño (PPT), el estimador del parámetro R es:

$$\hat{R}_{pppt} = \frac{\sum_{i=1}^n \frac{x_i}{p_i}}{\sum_{i=1}^n \frac{y_i}{p_i}}$$

Donde $p_i = \frac{L_i}{L}$ es la probabilidad de inclusión de la sección i . Siendo L_i = número de ciudadanos en lista nominal de la sección i y $L = 58,760,114$ el total de lista nominal en el padrón.

El error cuadrático medio de este estimador, usando la expresión 1.1, es:

$$ECM(\hat{R}) = \frac{Var(X_{pppt}) + R^2 Var(Y_{pppt}) - 2R cov(X_{pppt}, Y_{pppt})}{Y^2} \quad (1.3)$$

a) Con reemplazo

Si el muestreo se realiza con reemplazo, se tienen las siguientes expresiones para las varianzas y covarianza:

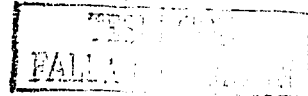
$$Var(X_{ppptcr}) = \frac{\sum_{i=1}^N p_i \left(\frac{X_i}{p_i} - X \right)^2}{n}$$

$$Var(Y_{ppptcr}) = \frac{\sum_{i=1}^N p_i \left(\frac{Y_i}{p_i} - Y \right)^2}{n}$$

$$Cov(X_{ppptcr}, Y_{ppptcr}) = \frac{\sum_{i=1}^N p_i \left(\frac{X_i}{p_i} - X \right) \left(\frac{Y_i}{p_i} - Y \right)}{n}$$

Donde $X = \sum_{i=1}^N X_i$, $Y = \sum_{i=1}^N Y_i$ son los totales poblacionales de cada variable.

Sustituyendo estas expresiones en la ecuación 1.3 y reduciendo términos se tiene la expresión para el error cuadrático medio de este estimador:



$$ECM(\hat{R}_{ppter}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{p_i} (X_i - RY_i)^2 \quad (1.4)$$

(Cochran, 1993, pág. 312 ec. 9A.8)

b) Sin reemplazo

En el caso del muestreo PPT sin reemplazo, Hartley y Rao (1962) proponen un método para la asignación de probabilidades de selección de las unidades en muestra, deduciendo las siguientes expresiones aproximadas para las varianzas y la covarianza:

$$Var(X_{pptsr}) = \frac{\sum_{i=1}^N p_i [1 - (n-1)p_i] \left(\frac{X_i}{p_i} - X \right)^2}{n}$$

$$Var(Y_{pptsr}) = \frac{\sum_{i=1}^N p_i [1 - (n-1)p_i] \left(\frac{Y_i}{p_i} - Y \right)^2}{n}$$

$$Cov(X_{pptsr}, Y_{pptsr}) = \frac{\sum_{i=1}^N p_i [1 - (n-1)p_i] \left(\frac{X_i}{p_i} - X \right) \left(\frac{Y_i}{p_i} - Y \right)}{n}$$

Sustituyendo estas expresiones en 1.3 la expresión para el error cuadrático medio en el muestreo con probabilidad proporcional al tamaño sin reemplazo es:

$$ECM(\hat{R}_{pptsr}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{p_i} [1 - (n-1)p_i] (X_i - RY_i)^2 \quad (1.5)$$

(Hartley y Rao, 1962)

1.4 Muestreo estratificado

Si se utiliza el diseño de muestreo estratificado, existen al menos dos estimadores diferentes para la razón poblacional: a) el primero denominado estimador simple o separado que consiste en obtener estimadores de razón en cada estrato y formar posteriormente el estimador estratificado que aglutina las estimaciones en cada estrato mediante el método usual en muestreo estratificado; b) el segundo denominado estimador combinado, consiste en realizar estimaciones de la variable

de estudio y la variable auxiliar, por separado considerando el muestreo estratificado.

1.4.1 Estimación separada

Se parte del supuesto de que la población esta dividida en H estratos. En cada estrato $h = 1, 2, \dots, H$ se calcula el estimador de razón:

$$\hat{R}_h = \frac{\sum_{i=1}^{n_h} x_i}{\sum_{i=1}^{n_h} y_i}$$

donde n_h es el tamaño de muestra en el estrato h y $n = \sum_{h=1}^H n_h$ es la muestra total considerada.

El estimador de razón separado de R tiene la siguiente expresión:

$$\hat{R}_s = \sum_{h=1}^H W_h \hat{R}_h \quad \text{donde } W_h = \frac{N_h}{N}, \quad N_h = \text{población en el estrato } h, \quad N = \sum_{h=1}^H N_h$$

Para el error cuadrático medio de este estimador se tiene:

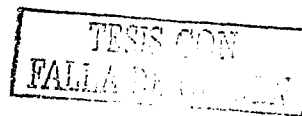
$$ECM(\hat{R}_s) = ECM\left(\sum_{h=1}^H W_h \hat{R}_h\right) = \sum_{h=1}^H ECM(W_h \hat{R}_h) = \sum_{h=1}^H W_h^2 ECM(\hat{R}_h)$$

Como en el estrato h se aplica el muestreo aleatorio simple sin reemplazo, sustituimos el error cuadrático medio correspondiente de cada \hat{R}_h , de acuerdo a la ecuación 1.2 con lo cual se obtiene la expresión:

$$ECM(\hat{R}_s) = \frac{N^2}{\gamma^2} \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \left(\frac{1}{N_h - 1}\right) \sum_{i=1}^{N_h} (X_{hi} - R_h Y_{hi})^2 \quad (1.6)$$

(Cochran, 1993, pág. 211, ec. 6.45)

En este diseño el sesgo de \hat{R}_s puede no ser despreciable en relación a su error estándar cuando los n_h son pequeños y el número de estratos H es grande, lo cual traería por consecuencia una acumulación del sesgo. (Cochran, 1993, pág. 212).



1.4.2 Estimación combinada

Para el estimador combinado, se considera la estimación de la media para las dos variables y el cociente de estos estimadores es el estimador de razón. Es decir:

$$\hat{R}_C = \frac{\bar{x}_{st}}{\bar{y}_{st}} = \frac{\sum_{h=1}^H N_h \bar{x}_h}{\sum_{h=1}^H N_h \bar{y}_h} \quad \text{donde} \quad \bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}, \quad \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

El error cuadrático medio se puede obtener, usando la expresión 1.1 de la siguiente manera:

$$ECM(\hat{R}_C) = \frac{1}{\bar{Y}_2} \left(Var(\bar{x}_{st}) + R^2 Var(\bar{y}_{st}) - 2RCov(\bar{x}_{st}, \bar{y}_{st}) \right)$$

En el muestreo estratificado las expresiones para las varianzas de las medias y su covarianza son:

$$Var(\bar{x}_{st}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_{Xh}^2}{n_h}, \quad Var(\bar{y}_{st}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_{Yh}^2}{n_h}$$

$$\text{y } Cov(\bar{x}_{st}, \bar{y}_{st}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_{XYh}}{n_h}$$

Sustituyendo estas igualdades se tiene:

$$ECM(\hat{R}_C) = \frac{1}{\bar{Y}_2} \left(\sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_{Xh}^2}{n_h} + R^2 \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_{Yh}^2}{n_h} - 2R \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_{XYh}}{n_h} \right)$$

Agrupando términos, finalmente el error cuadrático medio del estimador combinado de razón esta dado por la expresión:

$$ECM(\hat{R}_C) = \frac{N^2}{Y^2} \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{1}{n_h} \frac{1}{N_h - 1} \sum_{i=1}^{N_h} [(X_{hi} - \bar{X}_h) - R(Y_{hi} - \bar{Y}_h)]^2 \quad (1.7)$$

(Cochran 1993 pág. 213, ec. 6.51)

1.5 Muestreo bietápico

En este diseño de muestreo se considera a la población agrupada en N conglomerados, llamados unidades primarias de muestreo, los cuales son

homogéneos entre sí y heterogéneos a su interior de manera que el tomar una muestra de ellos nos proporcione información sobre la población. Con el fin de ahorrar tiempo y dinero para cada conglomerado en muestra se selecciona una muestra de unidades, en este caso unidades secundarias de muestreo, para que constituyan la muestra final. Para fijar ideas, consideremos que en la primera etapa se consideran n conglomerados o unidades primarias y los tamaños de cada conglomerado son M_i con $i = 1, 2, \dots, n$ y $M = \sum_{i=1}^N M_i$ es el tamaño de la población.

En la segunda etapa se selecciona una submuestra de tamaño m_i unidades secundarias de entre las M_i del conglomerado. Ambas selecciones de la muestra se realizan sin reemplazo. En la primera etapa, la muestra puede ser seleccionada usando muestreo aleatorio simple o bien con probabilidades proporcionales al tamaño. En la segunda etapa generalmente se realiza utilizando muestreo aleatorio simple. A continuación se presentan las expresiones en cada caso.

1.5.1 Muestreo aleatorio simple en ambas etapas

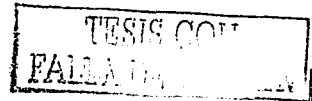
Si en ambas etapas del muestreo bietápico las unidades de muestreo se seleccionan mediante el muestreo aleatorio simple, el estimador de razón bajo este diseño esta dado por la expresión:

$$\hat{R}_{biemas} = \frac{\sum_{i=1}^n M_i \bar{x}_i}{\sum_{i=1}^n M_i \bar{y}_i}$$

donde $\bar{x}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij}$, $\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$ son las medias en muestra para cada variable, en cada uno de los conglomerados.

El error cuadrático medio de este estimador tiene la expresión:

$$ECM(\hat{R}_{biemas}) = \left(1 - \frac{n}{N}\right) \frac{N^2}{nY^2} \left(\frac{1}{N-1}\right) \sum_{i=1}^N (X_i - RY_i)^2 + \frac{N}{nY^2} \sum_{i=1}^N \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) S_{wi}^2 \quad (1.8)$$



$$\text{Donde } S_{wi}^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [x_{ij} - \bar{X}_i - R(y_{ij} - \bar{Y}_i)]^2$$

(Cochran, 1993, pág. 381, ec. 11.51)

1.5.2 Muestreo PPT en la primera etapa y muestreo aleatorio simple en la segunda etapa

Si en el muestreo bietápico las unidades primarias de muestreo se seleccionan con probabilidad proporcional al tamaño, dicha selección puede realizarse con o sin reemplazo. En este apartado se consideran las dos variantes.

a) Unidades primarias de muestreo seleccionadas con reemplazo

Si las unidades primarias de muestreo se seleccionan con probabilidad proporcional al tamaño con reemplazo y las unidades secundarias de muestreo se seleccionan mediante el muestreo aleatorio simple las expresiones para el estimador de razón y el error cuadrático medio de este estimador son:

$$\hat{R}_{biepptcr} = \frac{\sum_{i=1}^n \frac{M_i}{P_i} \bar{x}_i}{\sum_{i=1}^n \frac{M_i}{P_i} \bar{y}_i}$$

$$ECM(\hat{R}_{biepptcr}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} (X_i - RY_i)^2 + \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) S_{wi}^2 \quad (1.9)$$

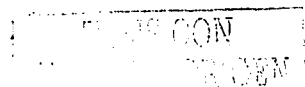
(Cochran, 1993, pag. 382, ec. 11.54)

Donde $p_i = \frac{M_i}{M}$ con $M = \sum_{i=1}^N M_i$ y con el cual estimador de razón es auto

ponderado. En este trabajo p_i puede definirse también como $p_i = \frac{L_i}{L}$ con L_i el número de individuos en lista nominal en el conglomerado i , L = total de lista nominal del padrón.

b) Unidades primarias de muestreo seleccionadas sin reemplazo

Si ahora las unidades primarias de muestreo se seleccionan de forma sistemática y con probabilidad proporcional al tamaño sin reemplazo y las unidades secundarias de muestreo se seleccionan mediante el muestreo aleatorio simple, Hartley y Rao



(1962) desarrollan un método para la asignación de probabilidades de selección de las unidades de primera etapa y deducen la siguiente expresión para el error cuadrático medio del estimador de razón.

$$ECM(\hat{R}_{bieppsr}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} [1 - (n-1)p_i] (X_i - RY_i)^2 + \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) S_{wi}^2 \quad (1.10)$$

(Hartley y Rao, 1962)

1.6 Sesgo y Coeficiente de Variación

El estimador de razón presentado en este trabajo es sesgado. En el caso del muestreo aleatorio simple, el sesgo puede calcularse por la expresión:

$$Sesgo(\hat{R}) = \frac{\left(1 - \frac{n}{N}\right)}{n\bar{Y}^2} (RS_Y^2 - S_{XY}) \quad (1.11)$$

(Cochran, 1993, pág. 208, ec. 6.33)

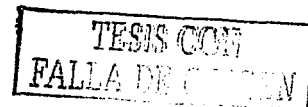
Por otra parte el coeficiente de variación de \bar{y} esta dado por:

$$cv(\bar{y}) = \frac{1}{\bar{Y}} \sqrt{\frac{\left(1 - \frac{n}{N}\right) S_y}{n}} \quad (1.12)$$

Se puede demostrar (Cochran 1993, pág. 209) que una cota para el sesgo relativo

esta dado por el coeficiente de variación de \bar{y} . Es decir: $\frac{sesgo(\hat{R})}{\sqrt{ECM(\hat{R})}} \leq CV(\bar{y})$. Un

criterio establecido por este autor es que el sesgo es despreciable si el coeficiente de variación es menor de 0.1.



Capítulo 2. Análisis Exploratorio del Listado Nominal de las Secciones Electorales

En este capítulo se presenta un análisis exploratorio de la composición del padrón electoral por cada sección, considerando variables como circunscripción, estado, distrito, y tipo de sección (urbana, rural o mixta). Asimismo se presenta un análisis descriptivo de los resultados obtenidos por los diferentes contendientes en las elecciones presidenciales efectuadas el 2 de julio del 2000.

2.1 Composición del padrón electoral

El listado nominal del padrón electoral estuvo integrado por 58,760,114 ciudadanos los cuales estuvieron distribuidos en 63,445 secciones electorales. El Instituto Federal Electoral (IFE) agrupó los 31 estados y el Distrito Federal (en este texto hablaremos de 32 estados) en 5 circunscripciones. La observación de la distribución del número de posibles votantes nos permite inferir que los criterios que se siguieron para conformar las circunscripciones son de tipo geográfico y con cierta tendencia a repartir equitativamente el número de personas inscritas en el padrón electoral (ver cuadro 2.1). En la figura 2.1 se puede observar que en términos porcentuales generales, las entidades que contemplan mayor cantidad de ciudadanos según lista nominal son: Estado de México (12.8%), Distrito Federal (10.6%), Veracruz (7.2%) y Jalisco (6.5%). En contraparte los estados con menor porcentaje son: Quintana Roo (0.8%), Campeche (0.7%), Colima (0.6%) y Baja California Sur (0.4%).

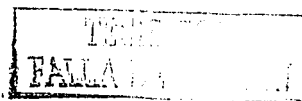
Otro tipo de agrupación es a nivel distrito, en este caso el IFE contempla un total de 300 distritos. El número de distritos por estado varía según el listado nominal del padrón estableciéndose una cierta relación proporcional entre estas dos variables. Por otra parte el número de distritos por circunscripción es muy parecido variando entre 59 y 61 (cuadro 2.1). Estas agrupaciones, la de circunscripciones y distritos, tienen que ver con la elección tanto directa, como de mayoría relativa de los 500 Diputados Federales y 128 Senadores de la república.



Con estos criterios el número de distritos por estado varía de acuerdo al número de ciudadanos empadronados, encontrando estados con un mínimo de dos distritos, Baja California Sur, Campeche, Colima y Quintana Roo, mientras que en el otro extremo existen entidades como el Distrito Federal con 30 distritos o el Estado de México con 36.

Los datos por sección electoral contempla una variable que define un criterio más de agrupación: si la sección es urbana, rural o mixta. En el cuadro 2.2 se presenta la distribución porcentual a nivel nacional de cada tipo de sección. Resalta que del total de secciones urbanas del país el 25% se encuentran en el Distrito Federal y el Estado de México. En contraparte, los estados con mayor porcentaje de casillas rurales son Oaxaca (11.4%), Veracruz (7.1%) y Puebla (6.9%). En el cuadro 2.3 y en la figura 2.2 se presentan la distribución de esta variable en cada estado. Se observa que el Distrito Federal, Baja California, Nuevo León y Coahuila son los que tienen un mayor porcentaje de población empadronada en secciones urbanas con el 98.2%, 88.1% y 84.2% respectivamente. En contraparte los estados con mayor población rural en la composición de su listado nominal son Chiapas(31.9%), Tabasco(46.8%), Hidalgo (25.3%) y Guerrero (29.8%).

Para efectos electorales la agrupación de secciones urbanas, rurales y mixtas resulta relevante para planear las encuestas debido a que la votación se manifiesta de forma diferenciada dependiendo si el tipo de población es urbana ó rural. Resultados de elecciones anteriores indican que el mayor porcentaje de votación rural es para el PRI mientras que la población de zona urbana, sobre todo de las grandes ciudades del interior de la república, han votado en los últimos años por el PAN, principal organización política de la coalición Alianza para el Cambio; en la Ciudad de México en las elecciones para jefe de gobierno de 1997 el triunfo correspondió al PRD principal partido político de la coalición Alianza por México.

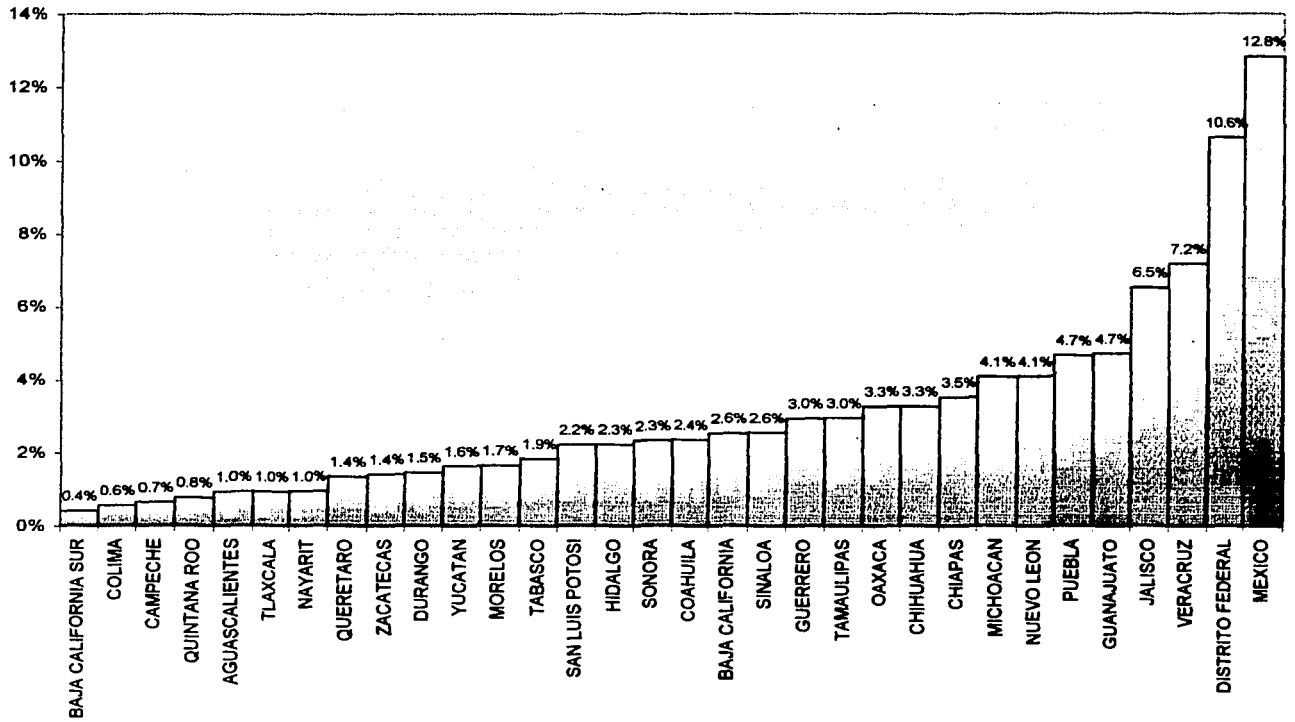


Cuadro 2.1. Distribución de secciones, lista nominal y distritos, por circunscripción y estado. Fuente IFE.

Estado	Secciones	% Secciones	Lista Nominal	% Lista Nominal	Distritos	% Distritos
BAJA CALIFORNIA	1,368	2.2	1,498,927	2.6	6	2.0
BAJA CALIFORNIA SUR	351	0.6	250,108	0.4	2	0.7
COLIMA	336	0.5	331,661	0.6	2	0.7
GUANAJUATO	3,004	4.7	2,784,469	4.7	15	5.0
JALISCO	3,306	5.2	3,845,518	6.5	19	6.3
NAYARIT	875	1.4	568,102	1.0	3	1.0
SINALOA	3,779	6.0	1,509,020	2.6	8	2.7
SONORA	1,326	2.1	1,378,548	2.3	7	2.3
Circunscripción 1	14,345	22.6	12,166,353	20.7	62	20.7
AGUASCALIENTES	486	0.8	562,390	1.0	3	1.0
CHIHUAHUA	2,691	4.2	1,937,458	3.3	9	3.0
COAHUILA	1,509	2.4	1,392,117	2.4	7	2.3
DURANGO	1,380	2.2	867,280	1.5	5	1.7
NUEVO LEON	2,102	3.3	2,412,803	4.1	11	3.7
QUERETARO	687	1.1	800,359	1.4	4	1.3
SAN LUIS POTOSI	1,791	2.8	1,315,650	2.2	7	2.3
TAMAULIPAS	1,734	2.7	1,746,782	3.0	8	2.7
ZACATECAS	1,872	3.0	838,034	1.4	5	1.7
Circunscripción 2	14,252	22.5	11,872,873	20.2	59	19.7
CAMPECHE	489	0.8	384,498	0.7	2	0.7
CHIAPAS	1,923	3.0	2,083,749	3.5	12	4.0
OAXACA	2,444	3.9	1,935,053	3.3	11	3.7
QUINTANA ROO	443	0.7	461,035	0.8	2	0.7
TABASCO	1,133	1.8	1,089,295	1.9	6	2.0
VERACRUZ	4,716	7.4	4,232,754	7.2	23	7.7
YUCATAN	1,057	1.7	969,074	1.6	5	1.7
Circunscripción 3	12,205	19.2	11,155,458	19.0	61	20.3
DISTRITO FEDERAL	5,533	8.7	6,256,316	10.6	30	10.0
HIDALGO	1,706	2.7	1,322,796	2.3	7	2.3
MORELOS	906	1.4	979,260	1.7	4	1.3
PUEBLA	2,548	4.0	2,753,111	4.7	15	5.0
TLAXCALA	607	1.0	562,589	1.0	3	1.0
Circunscripción 4	11,300	17.8	11,874,072	20.2	59	19.7
GUERRERO	2,760	4.4	1,738,419	3.0	10	3.3
MEXICO	5,910	9.3	7,544,110	12.8	36	12.0
MICHOACAN	2,673	4.2	2,408,829	4.1	13	4.3
Circunscripción 5	11,343	17.9	11,691,358	19.9	59	19.7
Total	63,445		58,760,114		300	

TESIS CON
DISEÑO ORIGINAL

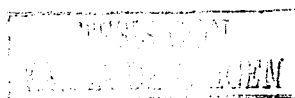
Figura 2.1. Porcentaje de electores registrados por estado según lista nominal



TESIS CON
FALSA

Cuadro 2.2. Distribución de las secciones electorales según tipo a nivel nacional.

Estado	Urbana	% Urbana	Rural	% Rural	Mixta	% Mixta	Total
AGUASCALIENTES	353	0.9	107	0.5	26	0.4	486
BAJA CALIFORNIA	1229	3.3	97	0.5	42	0.7	1368
BAJA CALIFORNIA SUR	218	0.6	91	0.5	42	0.7	351
CAMPECHE	260	0.7	184	0.9	45	0.8	489
CHIAPAS	614	1.6	1106	5.6	203	3.4	1923
CHIHUAHUA	1864	4.9	683	3.5	144	2.4	2691
COAHUILA	1084	2.9	260	1.3	165	2.8	1509
COLIMA	198	0.5	62	0.3	76	1.3	336
DISTRITO FEDERAL	5450	14.4	2	0.0	81	1.4	5533
DURANGO	569	1.5	736	3.7	75	1.3	1380
GUANAJUATO	1683	4.5	1171	5.9	150	2.5	3004
GUERRERO	823	2.2	1715	8.7	222	3.7	2760
HIDALGO	432	1.1	1099	5.6	175	2.9	1706
JALISCO	2156	5.7	751	3.8	399	6.7	3306
MEXICO	4338	11.5	1190	6.0	382	6.4	5910
MICHOACAN	1302	3.4	1032	5.2	339	5.7	2673
MORELOS	603	1.6	141	0.7	162	2.7	906
NAYARIT	464	1.2	343	1.7	68	1.1	875
NUEVO LEON	1693	4.5	274	1.4	135	2.3	2102
OAXACA	795	2.1	970	4.9	679	11.4	2444
PUEBLA	1323	3.5	812	4.1	413	6.9	2548
QUERETARO	348	0.9	268	1.4	71	1.2	687
QUINTANA ROO	288	0.8	97	0.5	58	1.0	443
SAN LUIS POTOSI	669	1.8	1011	5.1	111	1.9	1791
SINALOA	2397	6.3	1012	5.1	370	6.2	3779
SONORA	910	2.4	316	1.6	100	1.7	1326
TABASCO	477	1.3	597	3.0	59	1.0	1133
TAMAULIPAS	1193	3.2	369	1.9	172	2.9	1734
TLAXCALA	385	1.0	119	0.6	103	1.7	607
VERACRUZ	2353	6.2	1941	9.8	422	7.1	4716
YUCATAN	709	1.9	174	0.9	174	2.9	1057
ZACATECAS	584	1.5	995	5.0	293	4.9	1872
Total	37764		19725		5956		63445



Cuadro 2.3. Distribución de lista nominal por tipo en cada entidad.

Estado	Urbana	% Urbana	Rural	% Rural	Mixta	% Mixta	Total
AGUASCALIENTES	415005	73.8	107708	19.2	39677	7.1	562390
BAJA CALIFORNIA	1366114	91.1	83968	5.6	48845	3.3	1498927
BAJA CALIFORNIA SUR	173969	69.6	35057	14.0	41082	16.4	250108
CAMPECHE	247627	64.4	91493	23.8	45378	11.8	384498
CHIAPAS	828295	39.8	998057	47.9	257397	12.4	2083749
CHIHUAHUA	1607588	83.0	238679	12.3	91191	4.7	1937458
COAHUILA	1171605	84.2	86436	6.2	134076	9.6	1392117
COLIMA	207165	62.5	32607	9.8	91889	27.7	331661
DISTRITO FEDERAL	6142446	98.2	1388	0.0	112482	1.8	6256316
DURANGO	540529	62.3	277972	32.1	48779	5.6	867280
GUANAJUATO	1778561	63.9	852424	30.6	153484	5.5	2784469
GUERRERO	862960	49.6	663983	38.2	211476	12.2	1738419
HIDALGO	518521	39.2	596385	45.1	207890	15.7	1322796
JALISCO	2866919	74.6	502718	13.1	475881	12.4	3845518
MEXICO	6023103	79.8	967356	12.8	553651	7.3	7544110
MICHOACAN	1366831	56.7	667515	27.7	374483	15.5	2408829
MORELOS	679294	69.4	101995	10.4	197971	20.2	979260
NAYARIT	361072	63.6	151661	26.7	55369	9.7	568102
NUEVO LEON	2124871	88.1	128498	5.3	159434	6.6	2412803
OAXACA	781152	40.4	564868	29.2	589033	30.4	1935053
PUEBLA	1595249	57.9	668585	24.3	489277	17.8	2753111
QUERETARO	443321	55.4	264130	33.0	92908	11.6	800359
QUINTANA ROO	331336	71.9	63806	13.8	65893	14.3	461035
SAN LUIS POTOSI	734690	55.8	483628	36.8	97332	7.4	1315650
SINALOA	1070810	71.0	282416	18.7	155794	10.3	1509020
SONORA	1083881	78.6	196912	14.3	97755	7.1	1378548
TABASCO	512211	47.0	510158	46.8	66926	6.1	1089295
TAMAULIPAS	1354376	77.5	217143	12.4	175263	10.0	1746782
TLAXCALA	435383	77.4	31507	5.6	95699	17.0	562589
VERACRUZ	2356866	55.7	1442253	34.1	433635	10.2	4232754
YUCATAN	732436	75.6	75020	7.7	161618	16.7	969074
ZACATECAS	382199	45.6	284452	33.9	171383	20.5	838034
Total	41096385	69.9	11670778	19.9	5992951	10.2	58760114

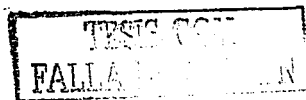
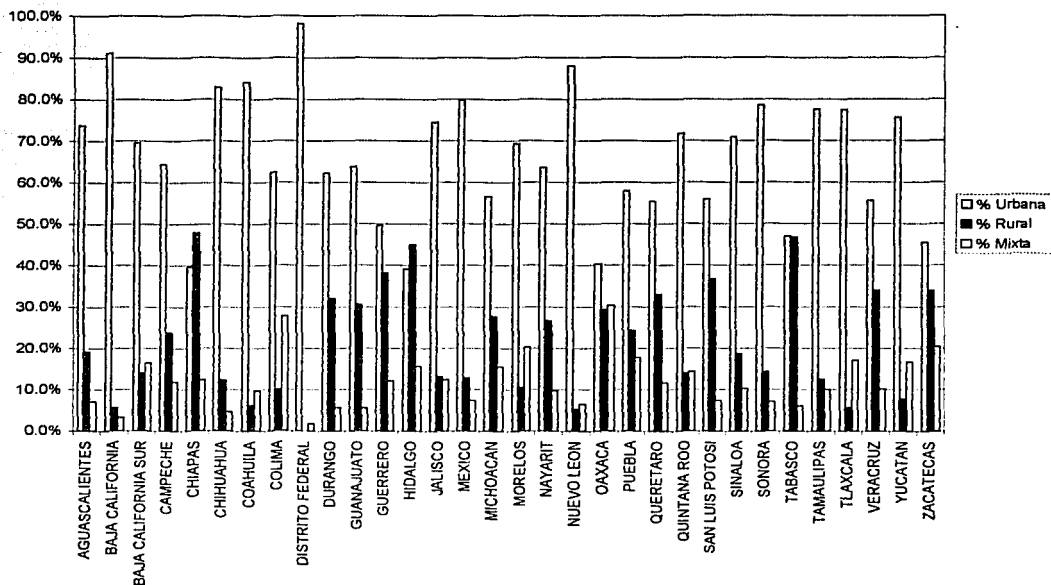


Figura 2.2. Tipo de sección electoral: porcentajes en cada entidad.



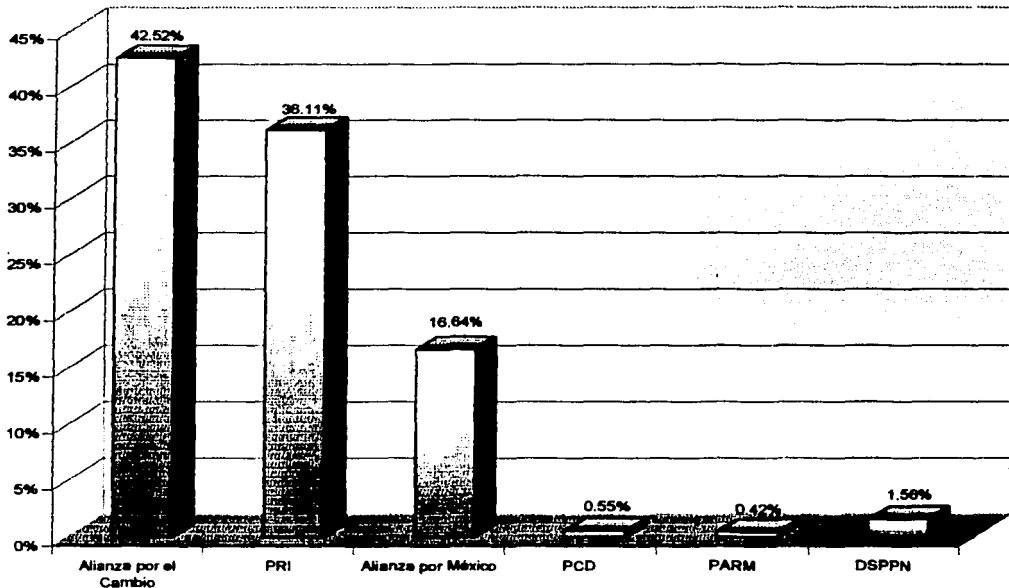
2.2 Resultado general de la elección

El día 2 de julio del 2000 se llevaron votaciones en toda la república para elegir Presidente de la República, Diputados Federales y Senadores y en el caso de algunos estados hubo elección de Gobernador, Presidentes Municipales, Jefe de Gobierno y Jefes Delegacionales en el Distrito Federal, y Diputados Locales. En este trabajo interesa sólo los resultados de la elección de Presidente de la República.

Para esta elección, de acuerdo a los datos del IFE, emitieron su voto un total de 37,601,618 personas, lo cual equivale al 63.99% del padrón total. Las preferencias obtenidas por cada partido a nivel nacional se manifestaron mayoritariamente por la coalición conformada por el PAN y el Partido Verde Ecologista de México denominada Alianza por el Cambio y que postuló a Vicente Fox Quezada con un 42.52%, mientras que el PRI que postuló a Francisco Labastida Ochoa quedó en segundo lugar con el

36.11%. Cuauhtémoc Cárdenas Solórzano postulado por la denominada Alianza por México, que estuvo conformada por 5 partidos que llevaba como cabeza al PRD, obtuvo el 16.64%. El PCD que postuló a Manuel Camacho Solís, el PARM cuyo candidato fue Porfirio Muñoz Ledo y DSPPN con Gilberto Rincón Gallardo tuvieron una votación menor al 2% lo que llevó a que esos partidos perdieran su registro ante el IFE. Estos resultados se presentan en la figura 2.3.

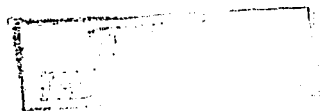
Figura 2.3. Resultados de las elecciones presidenciales 2000*



*Fuente: IFE, página electrónica: www.ife.org.mx

2.3 Resultados por circunscripción

Los resultados de la elección para los estados de la primera circunscripción que se presentan en el cuadro 2.4 y en la figura 2.4 nos indican que Alianza por el Cambio obtuvo el primer lugar en los estados de Baja California (49.76%), Colima (36.20%), Guanajuato (60.77%), Jalisco (53.07%), Sonora (50.79%) y en forma más cerrada en



Baja California Sur (36.20%). Por su parte el PRI ganó la elección en Nayarit (48.63%) y Sinaloa (64.00%). En todos los estados de esta circunscripción la Alianza por México obtuvo el tercer lugar siendo su votación porcentual más alta en Baja California Sur (26.91%) y más baja en Jalisco (6.22%). En la votación general de la circunscripción la Alianza por el Cambio obtuvo el 49.16% de los votos, el PRI 37.88% y la Alianza por México sólo el 8.80%.

Cuadro 2.4. Resultados de las elecciones: estados de la primera circunscripción

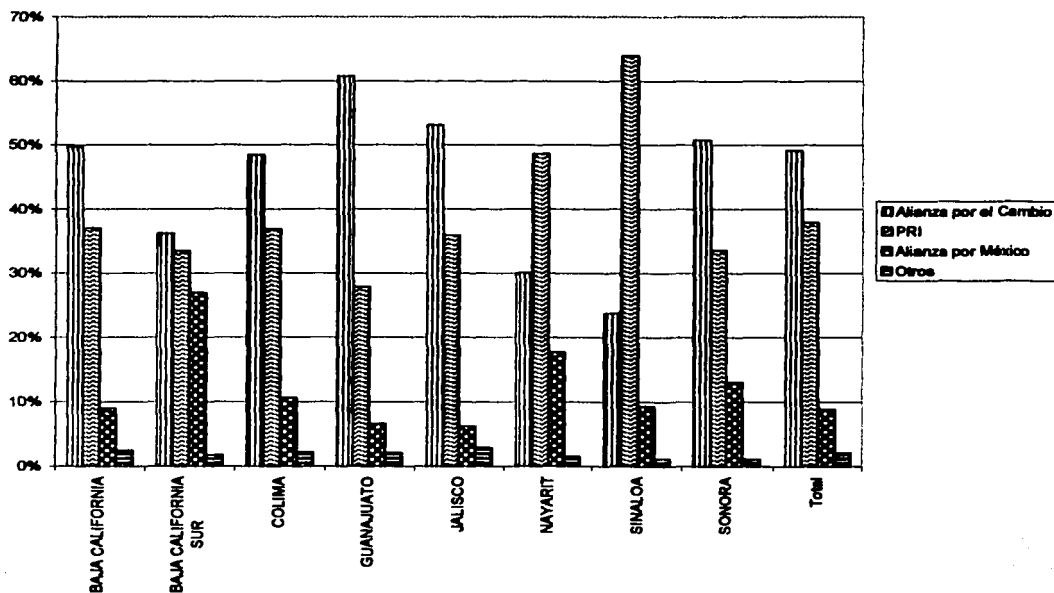
Primera Circunscripción: Resultados de Elección					
Estado	Alianza por el Cambio	PRI	Alianza por México	Otros	Total
BAJA CALIFORNIA	429194	319477	77340	21112	862595
%	49.76	37.04	8.97	2.45	
BAJA CALIFORNIA SUR	60834	56230	45229	2931	168045
%	36.20	33.46	26.91	1.74	
COLIMA	106445	81099	23313	4729	220002
%	48.38	36.86	10.60	2.15	
GUANAJUATO	1128780	517815	121489	37521	1857517
%	60.77	27.88	6.54	2.02	
JALISCO	1392535	941962	163269	74171	2623960
%	53.07	35.90	6.22	2.83	
NAYARIT	107417	173479	63121	5291	356702
%	30.11	48.63	17.70	1.48	
SINALOA	230777	621369	90488	11069	970913
%	23.77	64.00	9.32	1.14	
SONORA	447496	296267	114580	9423	881129
%	50.79	33.62	13.00	1.07	
Total	3903478	3007698	698829	166247	7940863
%	49.16	37.88	8.80	2.09	

En ocho de los nueve estados correspondientes a la segunda circunscripción se observó el mismo resultado Alianza por el Cambio obtuvo el primer lugar, el PRI el segundo y la Alianza por México el tercero. Como se puede observar en el cuadro 2.5 y la figura 2.5, los porcentajes para la Alianza por el Cambio fueron los siguientes: Aguascalientes(53.93%), Chihuahua (46.68%), Coahuila (48,87%), Durango (41.92%), Nuevo León (49.62%), Querétaro (51.94%), San Luis Potosí (47.44%), Tamaulipas (47.45%). El único estado con resultado diferente fue Zacatecas, donde el PRI obtuvo el 38.8%, seguido de la Alianza por el Cambio con 33.39% y la Alianza por México con

23.08%. Los resultados de toda la circunscripción ubicaron a la Alianza del Cambio en primer lugar con 47.75%, al PRI con 39.41% y a la Alianza por el Cambio con 8.82%.

La tercera circunscripción contempla los estados del sureste de México. El comportamiento de los electores en estos estados no tuvo el mismo comportamiento. El PRI quedó en primer lugar en Campeche (40.76%), Chiapas (43.09%), Oaxaca (42.73%) y Tabasco (39.45%). Por su parte la Alianza por el Cambio obtuvo el primer lugar en Quintana Roo (46.22%), Veracruz (39.88%) y Yucatán (47.10%). La Alianza por México obtuvo el primer lugar en Tabasco con el 39.45% de los votos, en los demás estados estuvo en tercer lugar. En toda la circunscripción el PRI obtuvo la mayoría de los votos (40.35%), seguido por la Alianza por el Cambio (35.08%) y la Alianza por México (20.11%). Los resultados completos se pueden observar en el cuadro 2.6 y la figura 2.6.

Figura 2.4 Porcentaje de votación para cada partido en los estados de la primera circunscripción.



Cuadro 2.5. Resultados de las elecciones: estados de la segunda circunscripción

Segunda Circunscripción: Resultados de Elección					
Estado	Alianza por el Cambio	PRI	Alianza por México	Otros	Total
AGUASCALIENTES	202335	127134	26264	13058	375165
%	53.93	33.89	7.00	3.48	
CHIHUAHUA	549177	460931	76810	19222	1128099
%	48.68	40.86	6.81	1.70	
COAHUILA	398800	311480	77393	14383	815974
%	48.87	38.17	9.48	1.76	
DURANGO	211361	222892	50592	9192	504190
%	41.92	44.21	10.03	1.82	
NUEVO LEON	760093	615907	96637	30584	1531941
%	49.62	40.20	6.31	2.00	
QUERETARO	290977	192622	39629	23023	560270
%	51.94	34.38	7.07	4.11	
SAN LUIS POTOSI	393997	324234	72599	16666	830576
%	47.44	39.04	8.74	2.01	
TAMAULIPAS	521486	445737	91426	19529	1098994
%	47.45	40.56	8.32	1.78	
ZACATECAS	169837	197336	117375	11178	508626
	33.39	38.80	23.08	2.20	
Total	3498063	2898273	648725	156835	7353835
	47.57	39.41	8.82	2.13	

En lo que se refiere a la cuarta circunscripción, el triunfo en las elecciones correspondió a la Alianza por el Cambio en tres entidades: Distrito Federal (43.65), Morelos (45.43) y Puebla (42.53), mientras que el PRI ganó en Hidalgo con el 43.50 y en Tlaxcala con el 36.44. La presencia más importante de la Alianza por México se dio en el Distrito Federal donde obtuvo el segundo lugar con 25.95 de los votos. En los otros estados ocupó el tercer lugar. Los resultados del total de la circunscripción indican que la Alianza por el Cambio obtuvo el más alto porcentaje con 42.42, seguido por el PRI con 30.53 y la Alianza por México con 21.33. En el cuadro 2.7 y figura 2.7 se presenta el total de resultados de esta circunscripción.

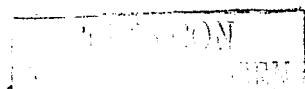
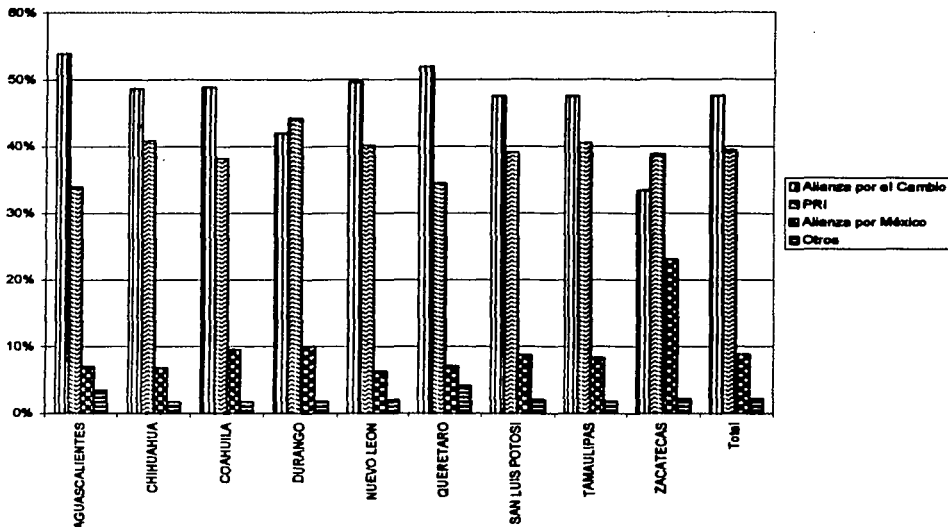
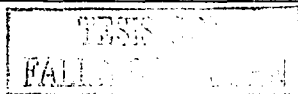


Figura 2.5 Porcentaje de votación para cada partido en los estados de la segunda circunscripción.



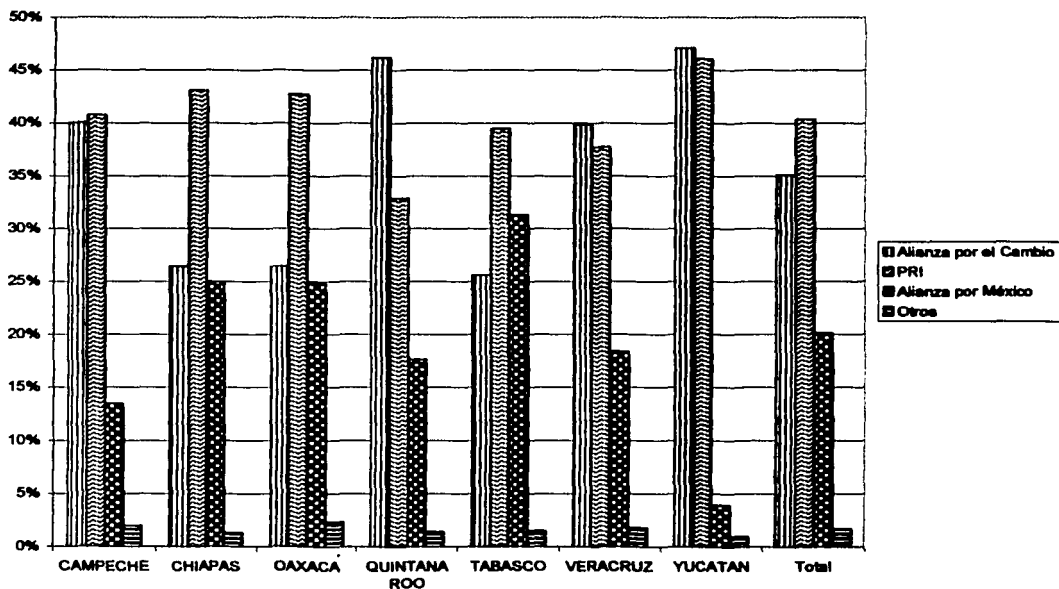
Cuadro 2.6. Resultados de las elecciones: estados de la tercera circunscripción

Tercera Circunscripción: Resultados de Elección					
Estado	Alianza por el Cambio	PRI	Alianza por México	Otros	Total
CAMPECHE	104498	106347	35090	5138	260941
%	40.05	40.76	13.45	1.97	
CHIAPAS	288204	469392	272182	14062	1089447
%	26.45	43.09	24.98	1.29	
OAXACA	301195	486496	282587	26751	1138496
%	26.46	42.73	24.82	2.35	
QUINTANA ROO	132383	94202	50487	4044	286402
%	46.22	32.89	17.63	1.41	
TABASCO	174840	269519	213983	10148	683181
%	25.59	39.45	31.32	1.49	
VERACRUZ	1066719	1008933	491791	47773	2674831
%	39.88	37.72	18.39	1.79	
YUCATAN	328503	321392	27214	6589	697427
%	47.10	46.08	3.90	0.94	
Total	4792687	5512565	2746669	229010	13661457
%	35.08	40.35	20.11	1.68	



Por último, en el cuadro 2.8 y la figura 2.8 se presentan los resultados de la quinta circunscripción que contempla sólo tres estados. Cada una de las tres principales organizaciones políticas contendientes obtuvo un primer lugar: Michoacán fue el único estado del país ganado por la Alianza por México con un porcentaje del 37.14, seguido por el PRI con 30.18. En el estado de México el triunfo correspondió a la Alianza por el Cambio con 43.71, siendo el segundo lugar en las preferencias el PRI con el 31.96. En Guerrero el PRI se llevó la victoria con el 42.67 de los votos seguido de la Alianza por México con 35.24. En términos globales, en esta circunscripción, la Alianza por el cambio obtuvo el 37.63, el PRI el 32.95 y la Alianza por México el 24.40.

Figura 2.6 Porcentaje de votación para cada partido en los estados de la tercera circunscripción

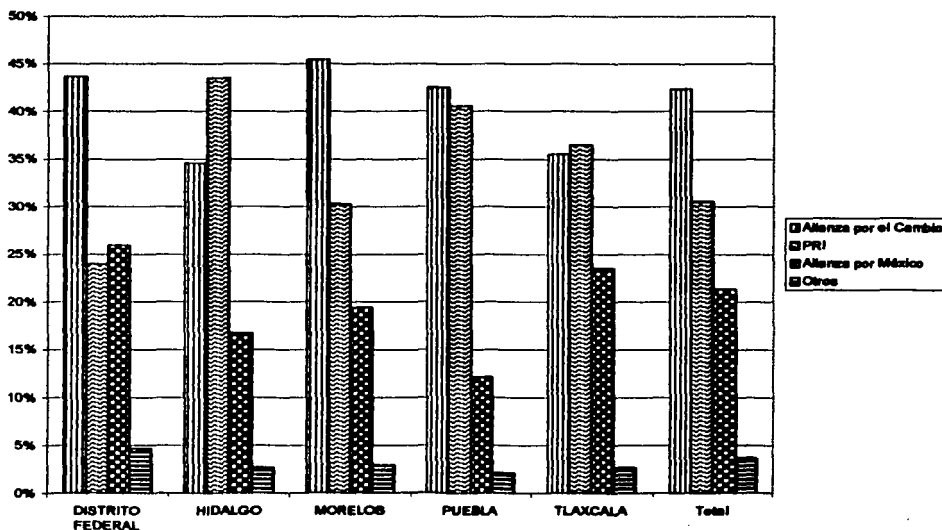


VESTIGIO DE
FALLA DE ORIGEN

Cuadro 2.7. Resultados de las elecciones: estados de la cuarta circunscripción

Cuarta Circunscripción: Resultados de Elección					
Estado	Alianza por el Cambio	PRI	Alianza por México	Otros	Total
DISTRITO FEDERAL	1928035 43.65	1060227 24.01	1146131 25.95	204538 4.63	4416609
HIDALGO	282864 34.60	355565 43.50	136861 16.74	21431 2.62	817476
MORELOS	290639 45.43	193861 30.30	124368 19.44	18465 2.89	639765
PUEBLA	732435 42.53	698974 40.59	208688 12.12	36628 2.13	1722172
TLAXCALA	123880 35.50	127163 36.44	82073 23.52	9143 2.62	348951
Total	6591828 42.42	4744418 30.53	3314170 21.33	571267 3.68	15540999 100.00

Figura 2.7 Porcentaje de votación para cada partido en los estados de la cuarta circunscripción



Cuadro 2.8. Resultados de las elecciones: estados de la quinta circunscripción

Quinta Circunscripción: Resultados de Elección					
Estado	Alianza por el Cambio	PRI	Alianza por México	Otros	Total
GUERRERO	174962	402091	332091	12095	942373
%	18.57	42.67	35.24	1.28	
MEXICO	2239750	1637714	961876	189073	5124572
%	43.71	31.96	18.77	3.69	
MICHOACAN	419188	441871	543804	26906	1464277
%	28.63	30.18	37.14	1.84	
Total	5667801	4963353	3675543	456148	15062447
	37.63	32.95	24.40	3.03	

En la figura 2.9 se presenta un mapa de la República Mexicana identificando cada entidad de acuerdo al partido que obtuvo la votación mayoritaria. Así se tiene marcado con azul los estados donde el triunfo correspondió a Alianza por el Cambio, con verde las entidades donde el triunfo fue para el PRI y con amarillo el caso del estado de Michoacán, única entidad ganada por la Alianza por México.

Figura 2.8 Porcentaje de votación para cada partido en los estados de la cuarta circunscripción

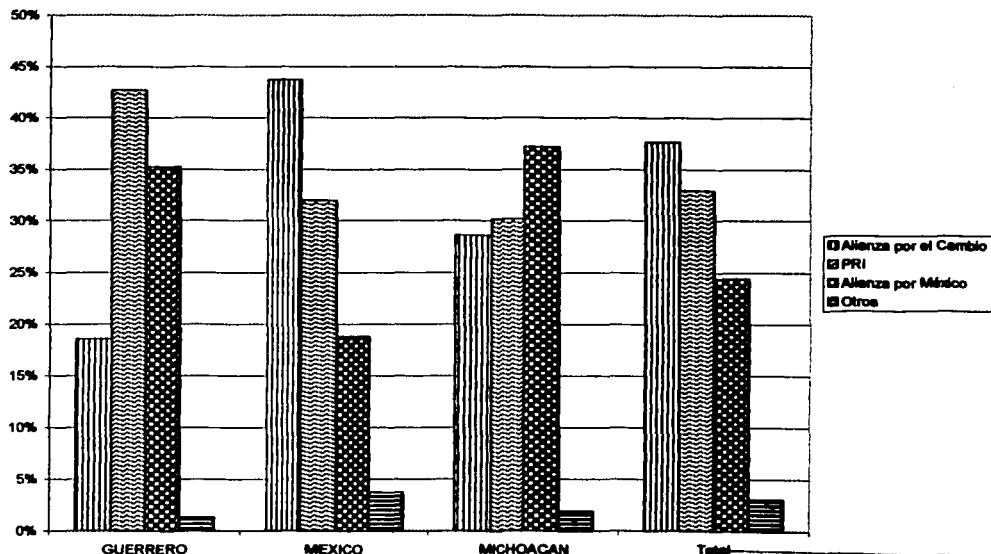
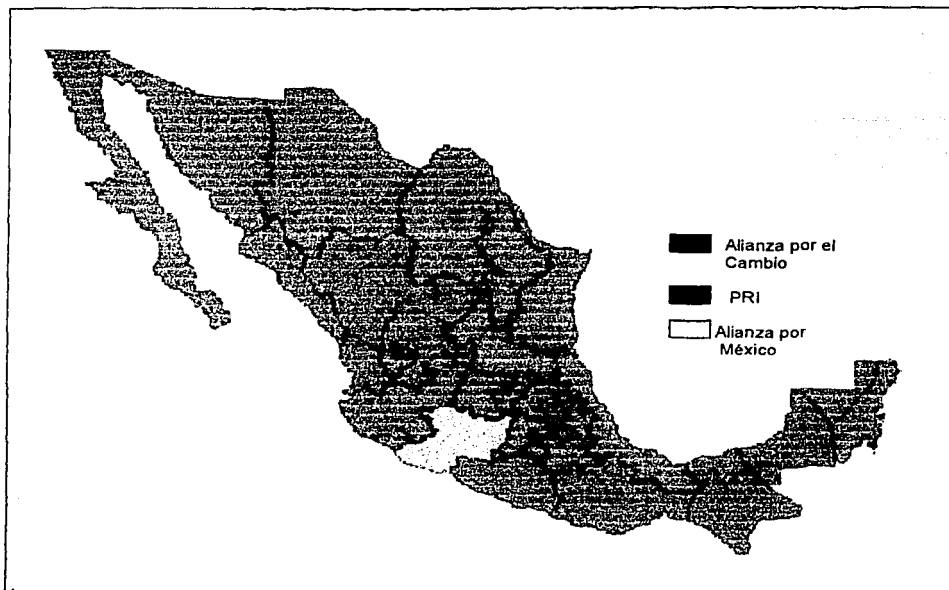


Figura 2.9 Entidades según voto mayoritario para los diferentes partidos políticos.

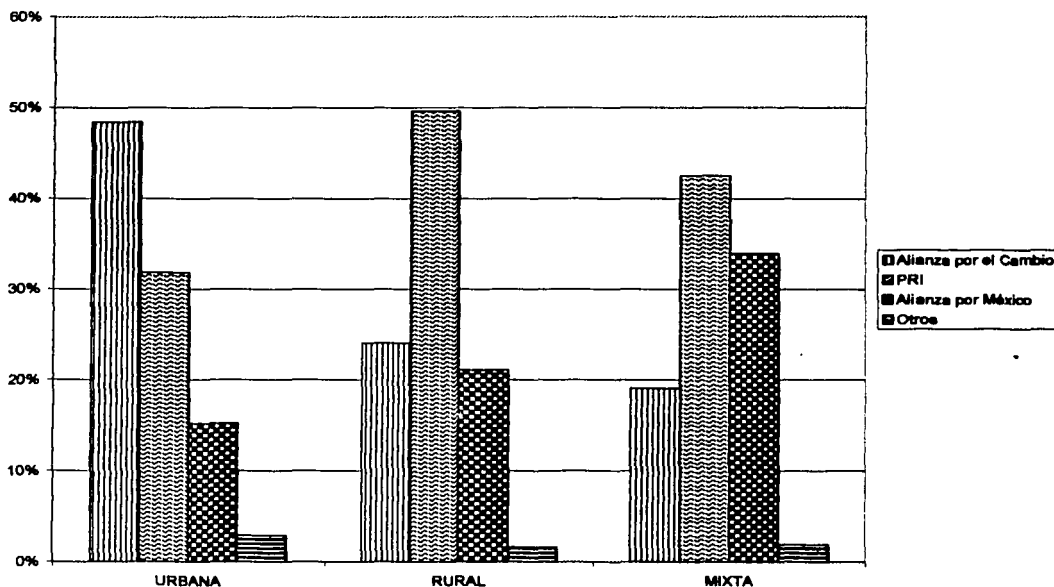


2.4 Resultados según tipo de sección

Ya se mencionó la importancia de considerar, para los estudios electorales, las secciones ubicadas en zonas rurales y urbanas. En relación a las elecciones presidenciales del año 2000, en la figura 2.10 se presentan los porcentajes obtenidos por cada partido, en cada uno de los estratos. Se puede observar que en las secciones urbanas la preferencia mayor se presentó en la opción Alianza por el Cambio (48.4), mientras que el PRI obtuvo el 31.8 y la Alianza por México 15.2. Por otra parte en el estrato de las secciones rurales el PRI obtiene las mayores preferencias con el 49.6, la Alianza por el Cambio 24.1 y la Alianza por México 21.1. Lo anterior nos confirma la asociación entre estas dos variables, en el sentido de que el PRI tiene una fuerte presencia en zonas rurales, los analistas políticos lo definen como el voto verde, mientras que la Alianza por el Cambio tiene una fuerte presencia en zonas urbanas. Con respecto a la Alianza por México se

tiene una fuerte presencia en zonas urbanas. Con respecto a la Alianza por México se observa una mayor presencia en zonas rurales que urbanas. Este comportamiento diferenciado de los electores nos indica que los diseños muestrales para elecciones presidenciales deberán considerar esta estratificación.

Figura 2.10 Distribución de votos según tipo de sección electoral.



2.5 Descripción multivariada por estado

Para saber si algunos estados se comportaron de manera similar en la preferencia de los electores, se calcularon las proporciones que los seis partidos políticos obtuvieron en cada estado. Con estas proporciones se aplicaron las técnicas de análisis de componentes principales y análisis de conglomerados considerando los 32 estados como casos y las proporciones obtenidas por cada partido como variables.

En el análisis de componentes principales las tres primeras componentes absorbieron el 91.9% de la varianza original; la primer componente esta asociado en forma directa con las proporciones que le corresponden a la Alianza por el Cambio y en

asocia principalmente, en forma directa, con los porcentajes obtenidos por la Alianza por México (ver cuadro 2.9).

Cuadro 2.9. Coeficientes de las componentes principales y las proporciones de los partidos por estado.

Partido	Componente		
	1	2	3
Alianza por el Cambio	.752	-.629	-.194
PRI	-.725	-.282	.606
PCD	.655	.561	.164
Alianza por México	-.356	.878	-.293
PARM	.611	.350	.611

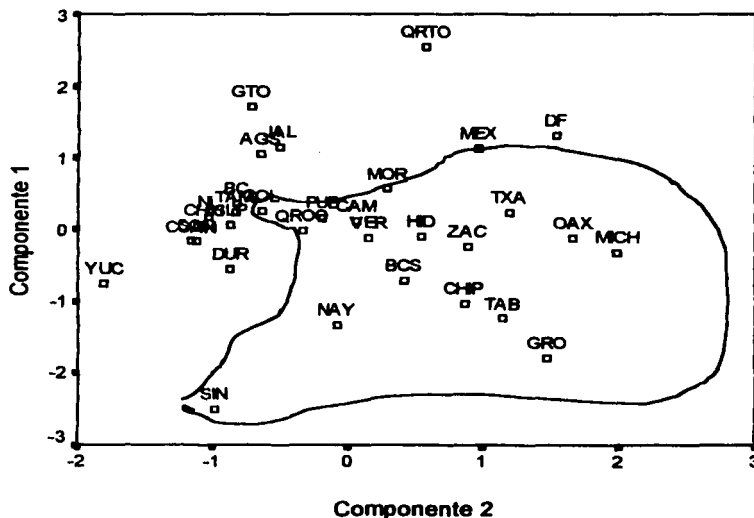
En la figura 2.11 se presenta la gráfica de las dos primeros componentes principales; se observa en la parte superior de la gráfica los estados en donde la Alianza por el Cambio obtuvo porcentajes más altos y en la parte inferior se ubican los estados donde el PRI obtuvo su mayor votación. En la parte derecha se ubican los estados donde la Alianza por México tuvo votación importante.

En el caso del Distrito Federal aunque la votación favoreció al candidato de la Alianza por el cambio, las proporciones alcanzadas por el PRI y la Alianza por México fueron similares por lo cual se comportan un tanto cuanto diferentes. Comportamiento atípico es también el de Yucatán donde los resultados estuvieron muy competidos entre la Alianza por el Cambio y el PRI, siendo casi nula la votación para las demás fuerzas políticas. En el caso de Sinaloa, es el estado donde un partido obtuvo la mayor proporción de votantes, correspondiendo al candidato del PRI quien es originario y fue gobernador de la entidad.

Al utilizar la técnica de análisis de conglomerados jerárquicos, con la distancia euclidiana y liga promedio, se construyó el dendograma que se presenta en la figura 2.12, donde se puede observar los dos grandes grupos de estados que se forman y tienen que ver, de manera similar a lo observado en el análisis de componentes principales, con las proporciones de votaciones para los dos principales partidos. El primer conglomerado, viendo de izquierda a derecha, está conformado por los estados donde la Alianza por el Cambio obtuvo una votación mayoritaria, mientras el segundo corresponde a estados donde el PRI obtuvo el primer lugar o su votación fue muy

importante, aunque no haya ganado, como es el caso de Michoacán. En la cuadro 2.10 se presentan las claves utilizadas para cada estado en el dendograma.

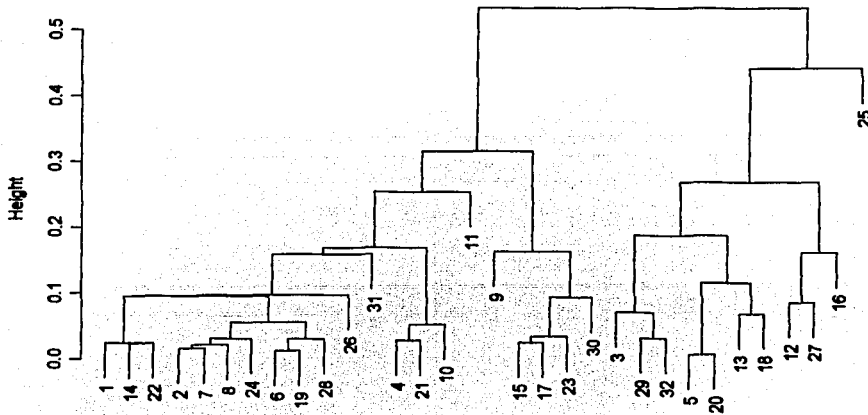
Figura 2.11. Gráfica de los dos primeros componentes principales, que explican el 71.6 de la varianza.



Cuadro 2.10. Claves por estado usados en el dendograma de la figura 2.11.

1 AGUASCALIENTES	9 DISTRITO FEDERAL	17 MORELOS	25 SINALOA
2 BAJA CALIFORNIA	10 DURANGO	18 NAYARIT	26 SONORA
3 BAJA CALIFORNIA SUR	11 GUANAJUATO	19 NUEVO LEON	27 TABASCO
4 CAMPECHE	12 GUERRERO	20 OAXACA	28 TAMAULIPAS
5 CHIAPAS	13 HIDALGO	21 PUEBLA	29 TLAXCALA
6 CHIHUAHUA	14 JALISCO	22 QUERETARO	30 VERACRUZ
7 COAHUILA	15 MEXICO	23 QUINTANA ROO	31 YUCATAN
8 COLIMA	16 MICHOACAN	24 SAN LUIS POTOSI	32 ZACATECAS

Figura 2.12. Dendograma que muestra la agrupación de los estados según la votación obtenida , utilizando análisis de conglomerados jerárquico.



SECRET
SALA DE TRABAJO

Capítulo 3: Resultados

En este capítulo se presentan los resultados de la evaluación del error cuadrático medio del estimador de razón para diferentes diseños de muestreo. Adicionalmente se analiza de forma separada el caso del muestreo bietápico.

3. 1 Notación

Los resultados que se presentan en este capítulo se obtuvieron al evaluar las expresiones del error cuadrático medio del estimador de razón en los diferentes diseños de muestreo mediante programas elaborados en Fortran (ver apéndice). En este apartado se especificará el significado de cada término en la expresión correspondiente del error cuadrático medio. Se respetará el número de ecuación dado en el capítulo 1. Al final del apartado se presentan las expresiones para el cálculo del sesgo del estimador de razón así como del coeficiente de variación de la media muestral en el caso del muestreo aleatorio simple.

3.1.1 Muestreo aleatorio simple

En el muestreo aleatorio simple la ecuación para el error cuadrático medio del estimador es:

$$ECM(\hat{R}) = \left(1 - \frac{n}{N}\right) \frac{N^2}{n} \left(\frac{1}{N-1}\right) \frac{1}{Y^2} \sum_{i=1}^N (X_i - RY_i)^2 \quad (1.2)$$

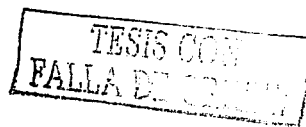
donde :

X_i = número de votos para el partido X en la sección electoral i .

Y_i = número total de votos emitidos en la sección electoral i .

$Y = \sum_{i=1}^N Y_i$ es el total de votos emitidos en la República Mexicana el 2 de julio del 2000.

N = número total de secciones electorales en todo el país. Para la elección del 2 de julio del 2000 $N = 63445$.



$R = \frac{\sum_{i=1}^N X_i}{\sum_{i=1}^N Y_i}$ es el parámetro de interés, el cual equivale a la proporción de

votos en la población.

n = número de secciones electorales muestreadas.

3.1.2 Muestreo con probabilidad proporcional al tamaño

Existen dos formas de calcular el error cuadrático medio en este diseño: la primera corresponde al caso cuando las unidades de muestreo se seleccionan con reemplazo y la segunda cuando es sin reemplazo. Las expresiones son:

a) Con reemplazo

$$ECM(\hat{R}_{pptr}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{p_i} (X_i - RY_i)^2 \quad (1.4)$$

Donde $p_i = \frac{L_i}{L}$ es la probabilidad de selección, siendo L_i el número de ciudadanos

en la lista nominal de la sección electoral i para $i=1, \dots, N$ y $L = \sum_{i=1}^N L_i = 58760114$.

b) Sin reemplazo

$$ECM(\hat{R}_{pptr}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{p_i} [1 - (n-1)p_i] (X_i - RY_i)^2 \quad (1.5)$$

3.1.4 Muestreo estratificado. Estimador separado

Si el diseño de muestreo es estratificado y se considera el estimador de razón separado, la ecuación para el error cuadrático medio es:

$$ECM(\hat{R}_s) = \frac{N^2}{Y^2} \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \left(\frac{1}{N_h - 1}\right) \sum_{i=1}^{N_h} (X_{hi} - R_h Y_{hi})^2 \quad (1.6)$$

donde:

H = número de estratos en que se divide la población. Se consideran cinco criterios de estratificación: circunscripciones, estados, distritos, circunscripción y tipo de sección, estado y tipo de sección.

N_h = número de secciones en el estrato h con $h=1, \dots, H$

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

n_h = número de secciones electorales muestreadas en el estrato h . La

asignación se realiza de manera proporcional, es decir $n_h = \frac{N_h}{N} n = W_h n$

X_{hi} = número de votos obtenidos por el partido X en la sección electoral i del estrato h , con $i=1, \dots, N_h$.

Y_{hi} = número de votos totales emitidos en la sección electoral i del estrato h , con $i=1, \dots, N_h$.

$$R_h = \frac{\sum_{i=1}^{N_h} X_{hi}}{\sum_{i=1}^{N_h} Y_{hi}}$$
 es la razón poblacional en el estrato h

3.1.5 Muestreo estratificado. Estimador combinado

Cuando el diseño de muestreo es estratificado y se considera el estimador de razón combinado, la ecuación para el error cuadrático medio es:

$$ECM(\hat{R}_c) = \frac{N^2}{Y^2} \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} \frac{1}{N_h - 1} \sum_{i=1}^{N_h} [(X_{hi} - \bar{X}_h) - R(Y_{hi} - \bar{Y}_h)]^2 \quad (1.7)$$

$$\bar{X}_h = \frac{\sum_{i=1}^{N_h} X_{hi}}{N_h}$$
 es el promedio poblacional de votos por sección electoral

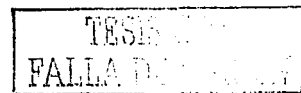
obtenido por el partido X en el estrato h .

$$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h}$$
 es el promedio poblacional de votos totales emitidos por sección

electoral en el estrato h .

3.1.6 Muestreo bietápico

En este diseño existen tres variantes: a) ambas etapas de muestreo se basan en el aleatorio simple; b) la primera etapa se toma con probabilidad proporcional al



tamaño con reemplazo y en la segunda se aplica el muestreo aleatorio simple; y c) en la primera se considera el muestreo con probabilidad proporcional al tamaño sin reemplazo y las unidades de segunda etapa con base en muestreo aleatorio simple. Las expresiones en cada caso son las siguientes:

a) Muestreo aleatorio simple en ambas etapas

En el muestreo bietápico considerando que en ambas etapas se aplica el muestreo aleatorio simple, el error cuadrático medio del estimador de razón es:

$$ECM(\hat{R}_{bmas}) = \left(1 - \frac{n}{N}\right) \frac{N^2}{nY^2} \left(\frac{1}{N-1}\right) \sum_{i=1}^N (X_i - RY_i)^2 + \frac{N}{nY^2} \sum_{i=1}^N \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) S_{wi}^2 \quad (1.8)$$

donde:

$$S_{wi}^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} [X_{ij} - \bar{X}_i - R(Y_{ij} - \bar{Y}_i)]^2$$

N = número de conglomerados o unidades primarias de muestreo. Se considera $N = 300$ distritos electorales.

n = es el número de distritos muestreados o unidades primarias de muestreo

M_i = número de secciones en el distrito i con $i = 1, 2, \dots, n$

m_i = número de secciones muestreadas o unidades secundarias de muestreo

X_{ij} = votos por el partido X en la sección j del distrito i

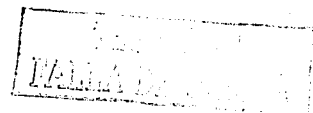
Y_{ij} = votos totales en la sección j del distrito i

X_i = total de votos para el partido X en el distrito i

Y_i = votos validos totales en el distrito i

$\bar{X}_i = \frac{\sum_{j=1}^{M_i} X_{ij}}{M_i}$ es la media de votos para el partido X en el distrito i

$\bar{Y}_i = \frac{\sum_{j=1}^{M_i} Y_{ij}}{M_i}$ es la media de votos totales en el distrito i



b) Muestreo con probabilidad proporcional al tamaño en la primera etapa con reemplazo y aleatorio simple en la segunda etapa

En el muestreo bietápico si las unidades primarias de muestreo se toman con probabilidad proporcional al tamaño en la primera etapa y se aplica muestreo aleatorio simple en la selección de unidades secundarias, el error cuadrático medio del estimador de razón es:

$$ECM(\hat{R}_{bieppter}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} (X_i - RY_i)^2 + \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) S_{wi}^2 \quad (1.9)$$

donde:

$p_i = \frac{\sum_{j=1}^{M_i} L_{ij}}{L}$ con L_{ij} = número de ciudadanos en la lista nominal de la sección j del distrito i , $L = 58760114$ es el total de ciudadanos en la lista nominal del padrón electoral. También puede definirse $p_i = \frac{M_i}{M}$ con $M = \sum_{i=1}^N M_i$ mediante lo cual el estimador de razón es auto ponderado.

c) Muestreo con probabilidad proporcional al tamaño en la primera etapa sin reemplazo y aleatorio simple en la segunda etapa

En este caso el error cuadrático medio tiene la expresión:

$$ECM(\hat{R}_{biepptsr}) = \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} [1 - (n-1)p_i] (X_i - RY_i)^2 + \frac{1}{nY^2} \sum_{i=1}^N \frac{1}{P_i} \frac{M_i^2}{m_i} \left(1 - \frac{m_i}{M_i}\right) S_{wi}^2 \quad (1.10)$$

de acuerdo al método de asignación de probabilidades propuesto por Hartley y Rao (1962).

3.1.8. Sesgo y coeficiente de variación en el muestreo aleatorio simple

Debido a que el estimador de razón es sesgado, es deseable conocer cuándo el sesgo es despreciable. En este sentido, un criterio dado por Cochran (pag. 208) es

que el sesgo será despreciable si se cumple la desigualdad $\frac{\text{sesgo}(\hat{R})}{\sqrt{\text{ECM}(\hat{R})}} \leq CV(\bar{y})$,

siendo ambas expresiones menores de 0.1. Se evalúa numéricamente este resultado en el caso del muestreo aleatorio simple donde el sesgo de \hat{R} tiene la expresión:

$$\text{Sesgo}(\hat{R}) = \frac{\left(1 - \frac{n}{N}\right)}{n\bar{Y}^2} (RS_Y^2 - S_{XY}) \quad (1.11)$$

y el coeficiente de variación de \bar{y} :

$$cv(\bar{y}) = \frac{1}{\bar{Y}} \sqrt{\frac{\left(1 - \frac{n}{N}\right) S_y}{n}} \quad (1.12)$$

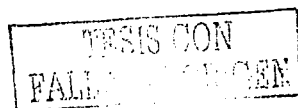
3.2 Error cuadrático medio

Se realizaron y ejecutaron programas en Fortran 90 (Anexo) para evaluar la estimación de la razón poblacional R para los diferentes diseños de muestreo con base en las ecuaciones 1.2, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 y 1.10 para el error cuadrático medio. Además se calcularon el sesgo y el coeficiente de variación para el muestreo aleatorio simple de acuerdo a las ecuaciones 1.11 y 1.12. En el cuadro 3.1 se presentan las siglas con que se identifican los 15 diseños de muestreo considerados, así como el número de ecuación correspondiente. Las secciones mixtas fueron agregadas a las secciones urbanas, por lo cual solo se hace alusión a dos tipos de secciones: urbanas y rurales.

En el cuadro 3.2 se presentan los valores calculados del Error Cuadrático Medio (ECM) en cada uno de los diseños de muestreo. Los valores están ordenados ascendentemente en cada una de las organizaciones políticas participantes. El número total de secciones electorales en muestra considerado es de 300, 600, 900, 1200, 1500 y 2000. El número que se presenta en el nombre de los diseños que se basan en muestreo estratificado esta relacionado con el número de estratos considerado.

Cuadro 3.1 Notación para los diseños de muestreo utilizados

Siglas	Diseño de muestreo	Ecuación
MAS	Muestreo Aleatorio Simple	1.2
BIEPPTCR	Muestreo bietápico, en la primera etapa se usa PPT con reemplazo y en la segunda MAS	1.9
BIEPPTSR	Muestreo bietápico, en la primera etapa se usa PPT sin reemplazo y en la segunda MAS	1.10
BIEMAS	Muestreo Bietápico, en ambas etapas se usa MAS	1.8
E5C	Muestreo estratificado con 5 circunscripciones, estimador combinado	1.7
E10C	Muestreo estratificado con 10 estratos debido a 5 circunscripciones por 2 tipos de sección (rural y urbana), estimador combinado	1.7
E300C	Muestreo estratificado con 300 distritos electorales, estimador combinado	1.7
E32C	Muestreo estratificado con 32 entidades, estimador combinado	1.7
E63C	Muestreo estratificado con 63 estratos debido a 32 entidades por 2 tipos de sección (rural y urbana), estimador combinado. En el DF se considero todas las secciones como urbanas.	1.7
E5S	Muestreo estratificado con 5 circunscripciones, estimador separado	1.6
E10S	Muestreo estratificado con 10 estratos debido a 5 circunscripciones por 2 tipos de sección (rural y urbana), estimador separado	1.6
E300S	Muestreo estratificado con 300 distritos electorales, estimador separado	1.6
E32S	Muestreo estratificado con 32 entidades, estimador separado	1.6
E63S	Muestreo estratificado con 63 estratos debido a 32 entidades por 2 tipos de sección (rural y urbana), estimador separado. En el DF se considero todas las secciones como urbanas.	1.6
PPTCR	Muestreo con probabilidad proporcional al tamaño con reemplazo	1.4
PPTSR	Muestreo con probabilidad proporcional al tamaño sin reemplazo	1.5



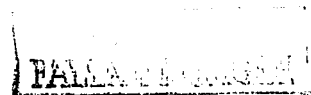
En el caso del muestreo bietápico se tomaron como unidades primarias de muestreo los distritos electorales y como unidades secundarias las secciones electorales y se presentan los resultados donde el valor del ECM es menor a partir de diferentes combinaciones que consideran 40, 60, 100, 150, 200 y 300 distritos, variando el número de secciones en cada distrito de 2 a 20, aclarando que para $n=2000$ se consideraron los 300 distritos en la primera etapa y 7 secciones en la segunda etapa de muestreo lo que da un total de $n=2100$ secciones electorales.

En la figuras 3.1, 3.2, 3.3 y 3.4 se presentan la gráficas de barras con el error de estimación, $\delta = 1.96\sqrt{ECM(\hat{R})}$ multiplicado por 100 para Alianza por el Cambio, PRI, Alianza por México y otros partidos respectivamente.

Los resultados numéricos y las gráficas muestran lo esperado: el ECM disminuye en relación al aumento de la muestra, en todos los diseños de muestreo. Los valores más grandes se presentan en el muestreo bietápico cuando las unidades de primera etapa se seleccionan con probabilidad proporcional al tamaño con reemplazo (BIEPPTCR), mientras que en el muestreo estratificado con 300 distritos, usando el estimador de razón separado (E300S), se obtuvieron los valores más pequeños.

En el cuadro 3.2 puede observarse que la mayor variabilidad se presenta en el estimador de la Alianza por el Cambio. Por otra parte a pesar de que los valores del ECM para el PRI son mayores respecto de los de la Alianza por México no aparentan ser significativamente diferentes. La variabilidad para otros partidos es muy baja, hecho que puede deberse a los porcentajes pequeños obtenidos en la contienda electoral.

Se puede ver que en los diseños que consideran algún tipo de estratificación el error cuadrático medio es menor al considerar el estimador separado respecto del combinado. En los diseños de una etapa basados en el muestreo con probabilidad proporcional al tamaño de la sección se observa que en el muestreo con reemplazo (PPTCR) el error es ligeramente mayor que sin reemplazo (PPTSR).



Cuadro 3.2. Error Cuadrático Medio para diferentes tamaños de muestra, considerando diferentes diseño de muestreo.

Muestra	Alianza por el Cambio		PRI		Alianza por México		Otros partidos	
	Diseño	$ECM(\hat{R}) \times 10^6$	Diseño	$ECM(\hat{R}) \times 10^6$	Diseño	$ECM(\hat{R}) \times 10^6$	Diseño	$ECM(\hat{R}) \times 10^6$
300	E300S	53.7	E300S	36.6	E300S	21.7	E300S	0.56
300	E300C	62.4	E300C	38.6	E300C	28.7	E300C	0.63
300	BIEPPTSR	66.8	BIEPPTSR	41.2	BIEPPTSR	30.5	E63S	0.67
300	E63S	68.9	E63S	42.3	E63S	33.3	BIEPPTSR	0.67
300	E63C	79.0	E63C	44.7	E32S	35.4	E63C	0.75
300	E10S	87.2	E10S	53.4	E63C	40.4	E32S	0.75
300	E32S	95.5	E10C	53.8	E32C	42.9	E32C	0.83
300	E10C	96.2	E32S	55.9	E10S	47.1	E10S	0.92
300	E32C	102.6	E32C	56.2	E5S	50.3	E10C	0.97
300	PPTSR	108.9	E5S	68.1	E10C	51.7	E5S	1.03
300	PPTCR	109.5	E5C	68.1	E5C	55.0	PPTSR	1.04
300	E5S	115.0	PPTSR	69.6	PPTSR	56.9	PPTCR	1.05
300	E5C	119.5	PPTCR	69.9	PPTCR	57.2	E5C	1.10
300	BIEMAS	124.3	MAS	73.5	BIEMAS	66.6	BIEMAS	1.20
300	MAS	129.4	BIEMAS	75.2	MAS	70.0	MAS	1.29
300	BIEPPTCR	191.4	BIEPPTCR	105.5	BIEPPTCR	109.0	BIEPPTCR	1.90
600	E300S	29.4	E300S	20.2	E300S	11.8	E300S	0.30
600	BIEMAS	32.5	BIEMAS	20.6	BIEMAS	15.1	BIEMAS	0.32
600	BIEPPTSR	33.4	BIEPPTSR	20.6	BIEPPTSR	15.3	BIEPPTSR	0.34
600	E63S	34.2	E63S	21.1	E300C	15.6	E300C	0.34
600	E300C	34.4	E300C	21.3	E63S	16.4	E63S	0.35
600	E63C	39.2	E63C	22.3	E32S	17.5	E32S	0.37
600	E10S	43.8	E10S	26.8	E63C	19.9	E63C	0.38
600	E32S	47.2	E10C	27.0	E32C	21.3	E32C	0.41
600	E10C	48.3	E32S	27.7	E10S	23.7	E10S	0.46
600	E32C	50.7	E32C	27.8	E5S	25.1	E10C	0.49
600	PPTSR	54.2	E5S	33.9	E10C	26.0	E5S	0.51
600	PPTCR	54.8	E5C	33.9	E5C	27.4	PPTSR	0.52
600	E5S	57.3	PPTSR	34.6	PPTSR	28.3	PPTCR	0.52
600	E5C	59.5	PPTCR	35.0	PPTCR	28.6	E5C	0.55
600	MAS	64.7	MAS	36.8	MAS	35.0	MAS	0.65
600	BIEPPTCR	95.7	BIEPPTCR	52.8	BIEPPTCR	54.5	BIEPPTCR	0.95
900	E300S	19.3	E300S	13.2	E300S	7.8	E300S	0.20
900	BIEMAS	21.5	BIEMAS	13.6	BIEMAS	10.0	BIEMAS	0.21
900	BIEPPTSR	22.1	BIEPPTSR	13.6	BIEPPTSR	10.1	BIEPPTSR	0.22

TESIS CON
 FALTA DE

Muestra	Alianza por el Cambio		PRI		Alianza por México		Otros partidos	
	Diseño	$ECM(\hat{R})x10^6$	Diseño	$ECM(\hat{R})x10^6$	Diseño	$ECM(\hat{R})x10^6$	Diseño	$ECM(\hat{R})x10^6$
900	E300C	22.5	E300C	13.9	E300C	10.3	E300C	0.22
900	E63S	22.7	E63S	14.0	E63S	10.9	E63S	0.23
900	E63C	26.0	E63C	14.7	E32S	11.6	E32S	0.24
900	E10S	29.0	E10S	17.8	E63C	13.2	E63C	0.25
900	E32S	31.2	E10C	17.9	E32C	14.1	E32C	0.27
900	E10C	32.0	E32S	18.3	E10S	15.7	E10S	0.31
900	E32C	33.5	E32C	18.4	E5S	16.7	E10C	0.32
900	PPTSR	35.9	E5S	22.5	E10C	17.2	E5S	0.34
900	PPTCR	36.5	E5C	22.5	E5C	18.2	PPTSR	0.34
900	E5S	38.1	PPTSR	23.0	PPTSR	18.8	PPTCR	0.35
900	E5C	39.5	PPTCR	23.3	PPTCR	19.1	E5C	0.36
900	MAS	43.1	MAS	24.5	MAS	23.3	MAS	0.43
900	BIEPPTCR	84.5	BIEPPTCR	45.8	BIEPPTCR	49.4	BIEPPTCR	0.84
1200	E300S	14.2	E300S	9.7	E300S	5.8	E300S	0.15
1200	BIEMAS	16.1	BIEPPTSR	10.2	BIEMAS	7.5	BIEMAS	0.16
1200	BIEPPTSR	16.5	BIEMAS	10.2	BIEPPTSR	7.6	BIEPPTSR	0.17
1200	E300C	16.6	E300C	10.3	E300C	7.6	E63S	0.17
1200	E63S	16.9	E63S	10.4	E63S	8.1	E300C	0.17
1200	E63C	19.3	E63C	11.0	E32S	8.7	E32S	0.18
1200	E10S	21.7	E10S	13.3	E63C	9.8	E63C	0.19
1200	E32S	23.3	E10C	13.4	E32C	10.5	E32C	0.20
1200	E10C	23.9	E32S	13.7	E10S	11.7	E10S	0.23
1200	E32C	25.1	E32C	13.8	E5S	12.4	E10C	0.24
1200	PPTSR	26.8	E5S	16.8	E10C	12.8	E5S	0.25
1200	PPTCR	27.4	E5C	16.8	E5C	13.6	PPTSR	0.26
1200	E5S	28.4	PPTSR	17.1	PPTSR	14.0	PPTCR	0.26
1200	E5C	29.5	PPTCR	17.5	PPTCR	14.3	E5C	0.27
1200	MAS	32.4	MAS	18.4	MAS	17.5	MAS	0.32
1200	BIEPPTCR	78.8	BIEPPTCR	42.4	BIEPPTCR	46.8	BIEPPTCR	0.78
1500	E300S	11.3	E300S	7.7	E300S	4.6	E300S	0.12
1500	BIEMAS	12.8	BIEPPTSR	8.1	BIEMAS	5.9	BIEMAS	0.13
1500	BIEPPTSR	13.1	BIEMAS	8.1	BIEPPTSR	6.0	BIEPPTSR	0.13
1500	E300C	13.2	E300C	8.1	E300C	6.0	E300C	0.13
1500	E63S	13.4	E63S	8.3	E63S	6.4	E63S	0.13
1500	E63C	15.4	E63C	8.7	E32S	6.9	E32S	0.15
1500	E10S	17.2	E10S	10.6	E63C	7.8	E63C	0.15

TESIS CON
 FALLA DE CALIDAD

Muestra	Alianza por el Cambio		PRI		Alianza por México		Otros partidos	
	Diseño	$ECM(\hat{R}) \times 10^6$	Diseño	$ECM(\hat{R}) \times 10^6$	Diseño	$ECM(\hat{R}) \times 10^6$	Diseño	$ECM(\hat{R}) \times 10^6$
1500	E32S	18.6	E10C	10.6	E32C	8.4	E32C	0.16
1500	E10C	19.0	E32S	10.9	E10S	9.3	E10S	0.18
1500	E32C	20.0	E32C	11.0	E5S	9.9	E10C	0.19
1500	PPTSR	21.3	E5S	13.4	E10C	10.2	E5S	0.20
1500	PPTCR	21.9	E5C	13.4	E5C	10.8	PPTSR	0.20
1500	E5S	22.6	PPTSR	13.6	PPTSR	11.1	PPTCR	0.21
1500	E5C	23.5	PPTCR	14.0	PPTCR	11.4	E5C	0.22
1500	MAS	25.9	MAS	14.7	MAS	14.0	MAS	0.26
1500	BIEPPTCR	75.5	BIEPPTCR	40.3	BIEPPTCR	45.3	BIEPPTCR	0.75
2000	E300S	8.5	E300S	5.7	E300S	3.4	E300S	0.09
2000	E300C	9.9	E300C	6.1	E300C	4.5	E300C	0.10
2000	E63S	10.0	E63S	6.2	E63S	4.8	E63S	0.10
2000	E63C	11.5	E63C	6.5	E32S	5.1	E32S	0.11
2000	E10S	12.8	E10S	7.9	E63C	5.8	E63C	0.11
2000	E32S	13.8	E10C	7.9	E32C	6.2	E32C	0.12
2000	E10C	14.2	E32S	8.1	E10S	6.9	E10S	0.14
2000	E32C	14.8	E32C	8.1	E5S	7.4	E10C	0.14
2000	PPTSR	15.8	E5S	9.9	E10C	7.6	E5S	0.15
2000	PPTCR	16.4	E5C	9.9	E5C	8.0	PPTSR	0.15
2000	E5S	16.8	PPTSR	10.1	PPTSR	8.3	PPTCR	0.16
2000	E5C	17.5	PPTCR	10.5	PPTCR	8.6	E5C	0.16
2000	MAS	19.4	MAS	11.0	MAS	10.5	MAS	0.19
2100	BIEMAS	9.0	BIEPPTSR	5.7	BIEMAS	4.2	BIEMAS	0.09
2100	BIEPPTSR	9.3	BIEMAS	5.7	BIEPPTSR	4.3	BIEPPTSR	0.09
2100	BIEPPTCR	71.6	BIEPPTCR	37.9	BIEPPTCR	43.5	BIEPPTCR	0.71

TESIS CON
FALLA DE CUMPLIMIENTO

Las expresiones matemáticas del ECM para los diseños E300S y BIEMAS (con 300 UPM) son iguales, sin embargo al asignar de manera proporcional la muestra en cada estrato se produce una muestra mayor, por lo cual numéricamente no coinciden.

Los valores del ECM en el muestreo bietápico basado en aleatorio simple en ambas etapas (BIEMAS) y con probabilidad proporcional sin reemplazo (BIEPPTSR) tomando una muestra de 300 distritos en la primera etapa ($n=600, 900, 1200$ y 1500) son similares al obtenido en el muestreo estratificado por distritos basado en el estimador combinado (E300C). La concepción de los dos muestreos (bietápico y estratificado) es diferente, debido a que mientras en el bietápico se parte de que las unidades de primera etapa (distritos) son homogéneas entre sí, en el estratificado deben ser heterogéneas. Considerar los 300 distritos en un muestreo en la práctica implicaría altos costos en el trabajo de campo, razón por la cual el diseño E300C podría no ser muy adecuado, siendo probablemente más viable proponer el bietápico con menos distritos en la primera etapa de muestreo.

En los resultados del ECM obtenidos por el PRI, se observa que para los diseños que consideran el muestreo estratificado, el ECM disminuye en relación al aumento de estratos. Se esperaría una situación similar en todos los casos, sin embargo resalta el caso de la Alianza por el Cambio donde el valor del ECM considerando la estratificación por entidad con el estimador separado (E32S) es mayor que el que contempla la estratificación por circunscripción y tipo (E10S); misma situación que se produce con el estimador combinado, lo que podría indicarnos mayor homogeneidad regional que contempla el tipo de sección rural respecto al comportamiento por entidad. Para la Alianza por México este patrón se observa al utilizar el estimador combinado, por ejemplo el error cuadrático medio para el diseño E32S es ligeramente mayor que el de E63C, a pesar de que el número de estratos es menor.

Para todos los partidos y los diferentes tamaños de muestra se cumple que si la estratificación se realiza por entidad y tipo (E63S) el ECM es mayor que el obtenido para E300S aunque no por mucho; por lo tanto, sería más viable en la

práctica debido a su menor costo en el campo al considerar sólo 63 estratos. La estratificación por circunscripción (E5S y E5C) producen errores de estimación ligeramente menores al muestreo aleatorio simple (MAS), por lo cual no es recomendable su implementación.

Aunque los errores cometidos en el muestreo proporcional al tamaño con reemplazo (PPTCR) y sin reemplazo (PPTSR) son menores que el cometido en el aleatorio simple, resultan ser mayores a los que consideran algún tipo de estratificación, salvo el caso de circunscripciones para la Alianza por el Cambio. La ventaja es que en estos diseños se puede bajar el costo al tomar secciones con mayor número de votantes.

Para comparar el comportamiento de los diseños basados en muestreo bietápico (BIEMAS, BIEPPTSR y BIEPPTCR) se calculó el ECM dejando constante el número de unidades secundarias de muestreo en 4 secciones y variando el número de unidades primarias de muestreo de 10 a 300 distritos. En las figuras 3.5, 3.6 y 3.7 se presentan las gráficas del error de estimación para los tres diseños para Alianza por el Cambio, PRI y Alianza por México respectivamente. Se puede observar que los valores más pequeños del error de estimación se obtienen en el diseño BIEPPTSR y los más grandes en el BIEPPTCR. Para que el muestreo BIEPPTCR presente valores similares a los otros diseños debería considerarse una muestra menor al 10% del total de distritos de manera que el efecto de diseño no afecte demasiado el valor del ECM. El muestreo bietápico con probabilidad proporcional al tamaño en la primera etapa sin reemplazo (BIEPPTSR) presenta valores similares al BIEMAS al aumentar el número de distritos en muestra, siendo más eficiente el BIEPPTSR con muestras más pequeñas, por lo cual puede ser un diseño adecuado para implementarse en la práctica.

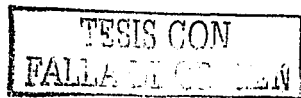
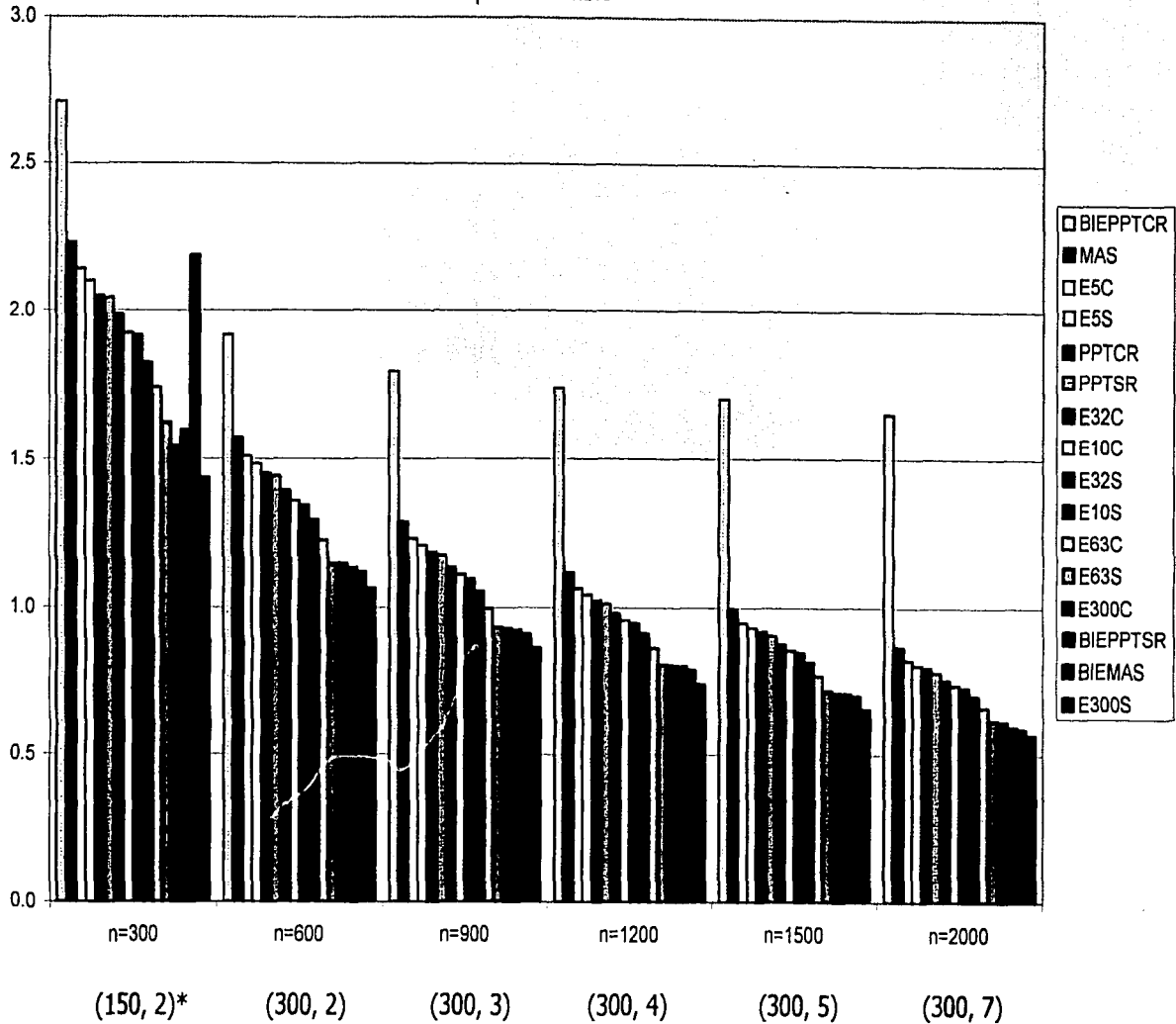
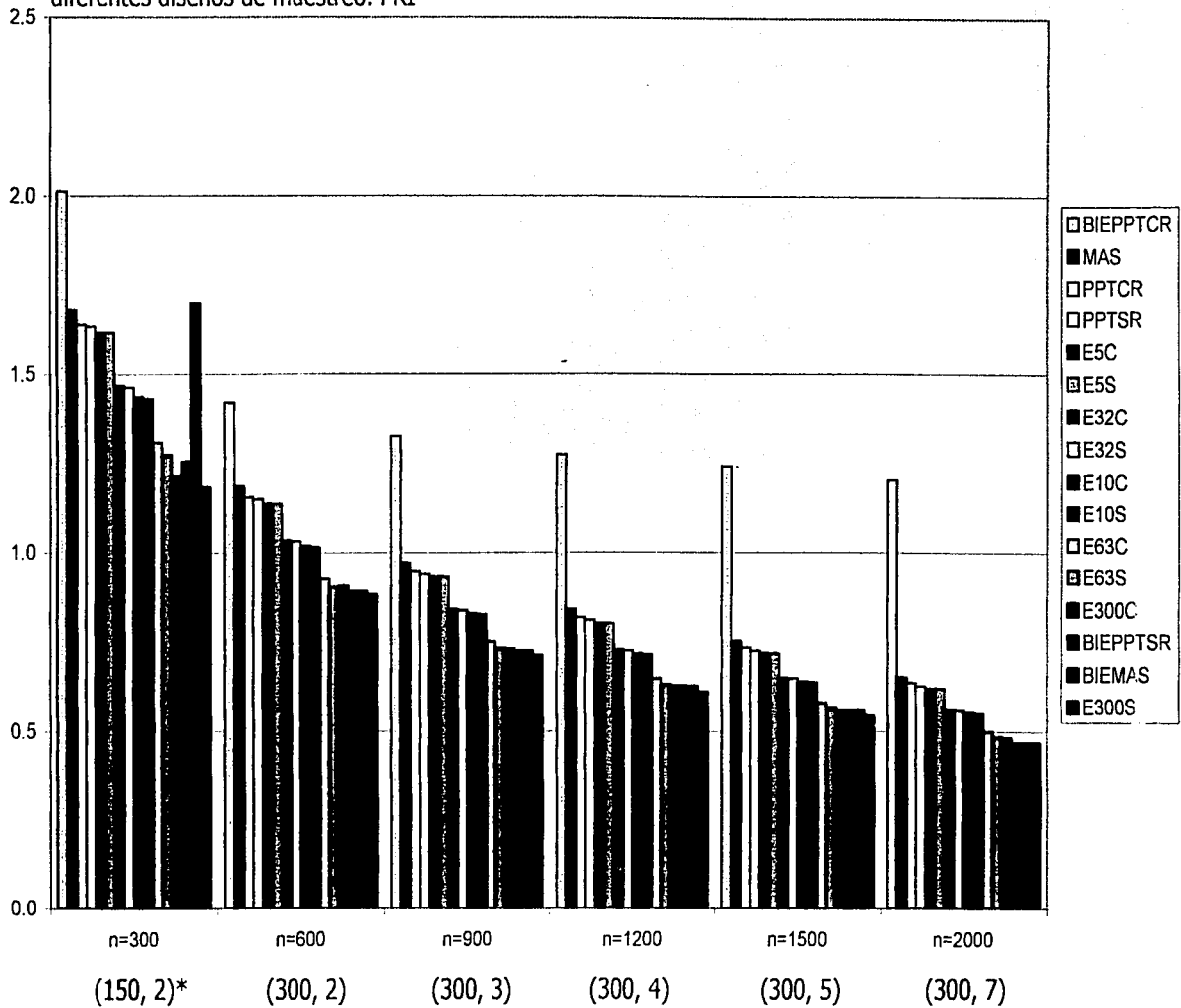


Figura 3.1. Error de estimación $\delta \times 10^2$ ($\delta = 1.96 \sqrt{ECM(\hat{R})}$) para varios tamaños de muestra, bajo diferentes diseños de muestreo. Alianza por el Cambio



* Las coordenadas indican UPM y USM en el muestreo bietápico: (UPM, USM)

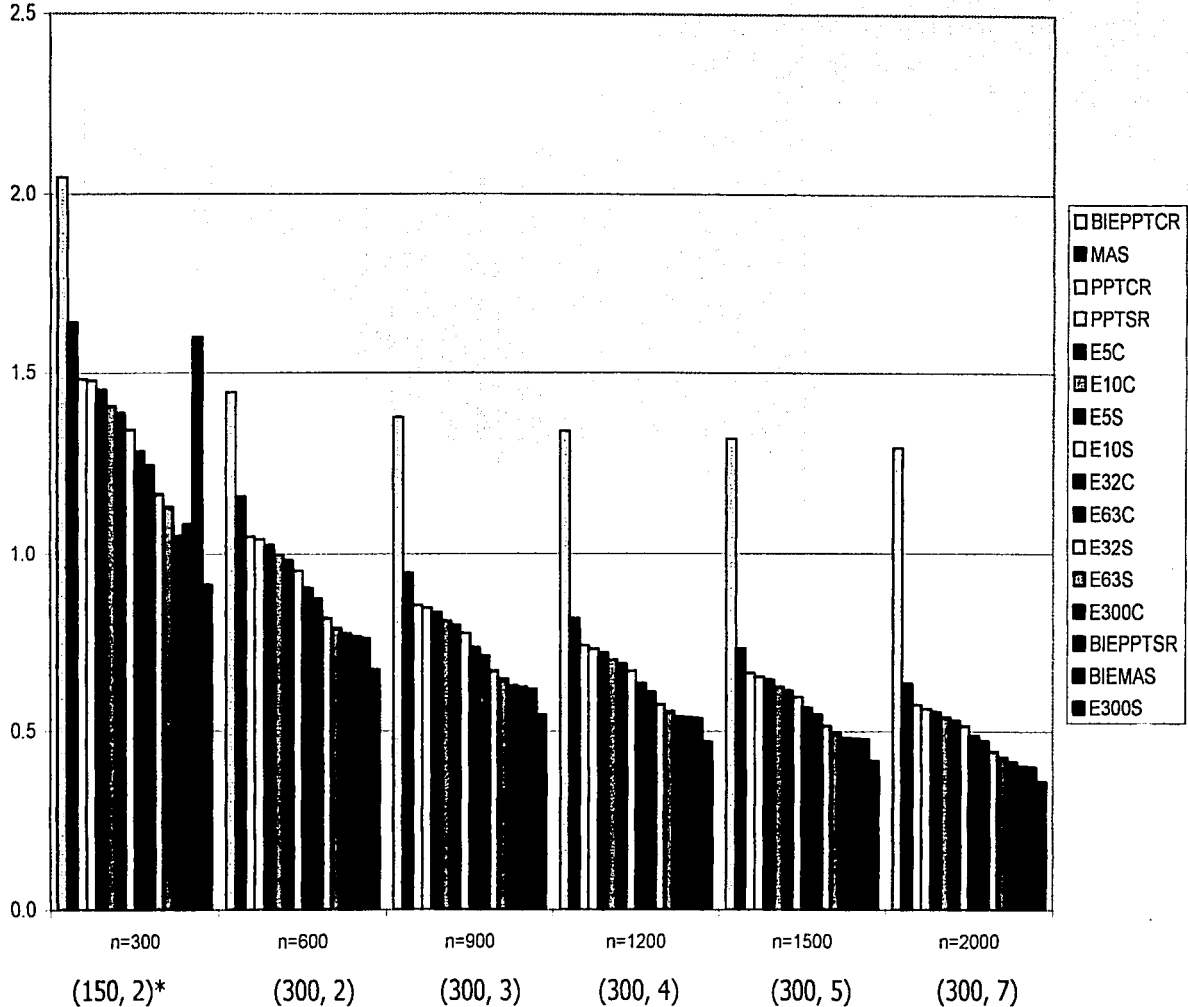
Figura 3.2. Error de estimación $\delta \times 10^2$ ($\delta = 1.96\sqrt{ECM(\hat{R})}$) para varios tamaños de muestra, bajo diferentes diseños de muestreo. PRI



* Las coordenadas indican UPM y USM en el muestreo bietápico: (UPM, USM)

INSTITUTO VENEZOLANO DE INVESTIGACIONES CIENTÍFICAS Y TECNOLÓGICAS CON FALTA DE ORIGEN

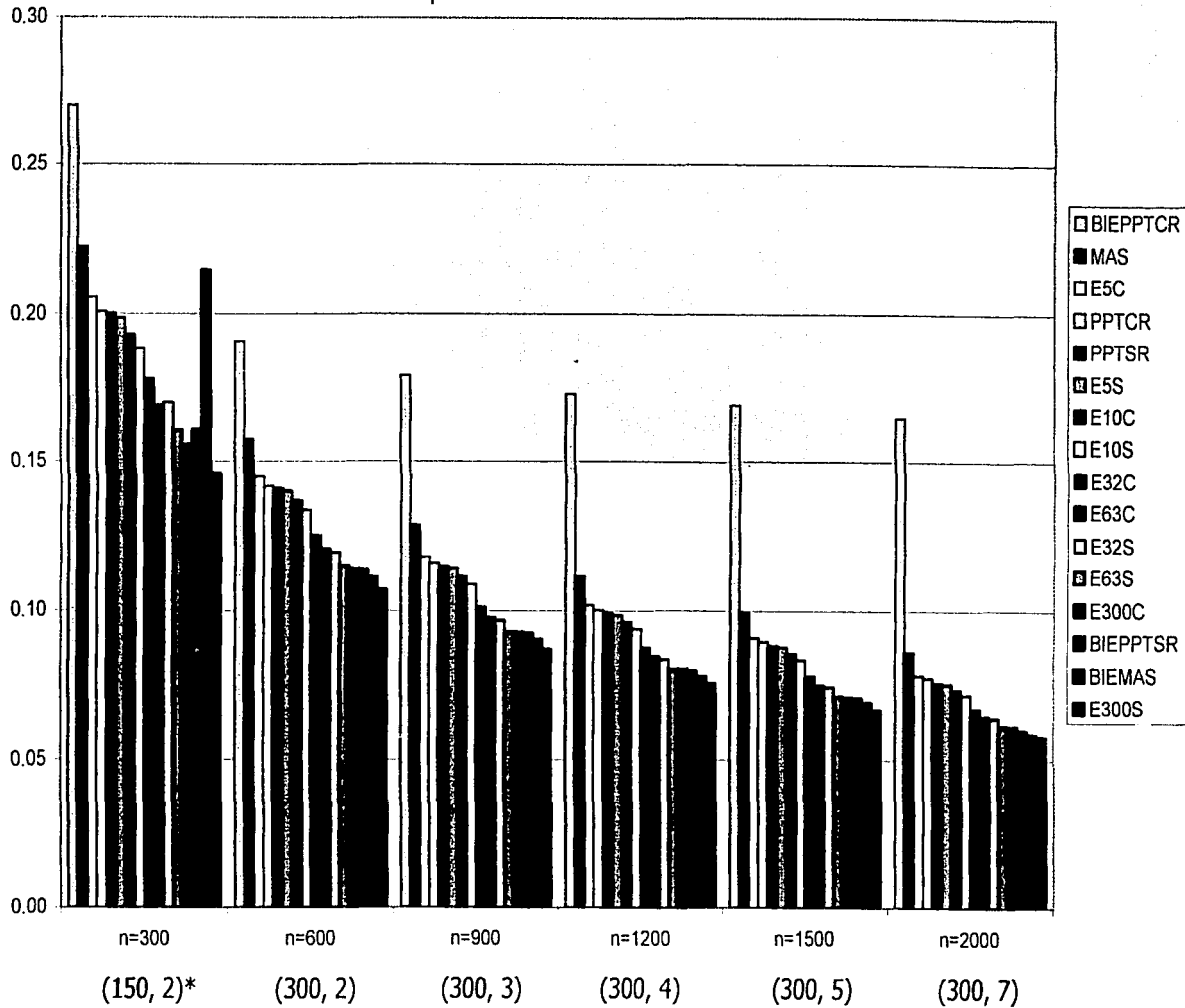
Figura 3.3. Error de estimación $\delta \times 10^2$ ($\delta = 1.96\sqrt{ECM(\hat{R})}$) para varios tamaños de muestra, bajo diferentes diseños de muestreo. Alianza por México



* Las coordenadas indican UPM y USM en el muestreo bietápico: (UPM, USM)

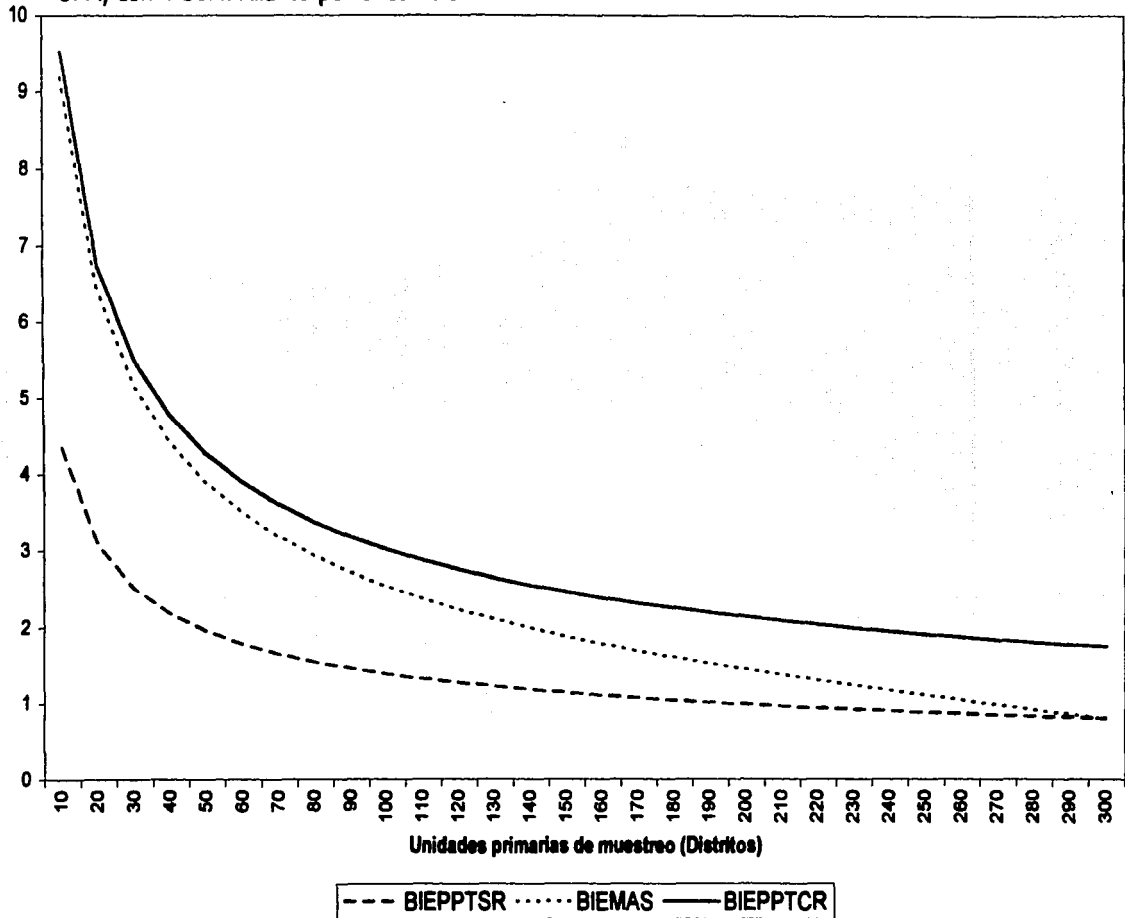
FALTA DE ORIGEN
 PARA COPIA

Figura 3.4. Error de estimación $\delta \times 10^2$ ($\delta = 1.96\sqrt{ECM(\hat{R})}$) para varios tamaños de muestra, bajo diferentes diseños de muestreo. Otros partidos



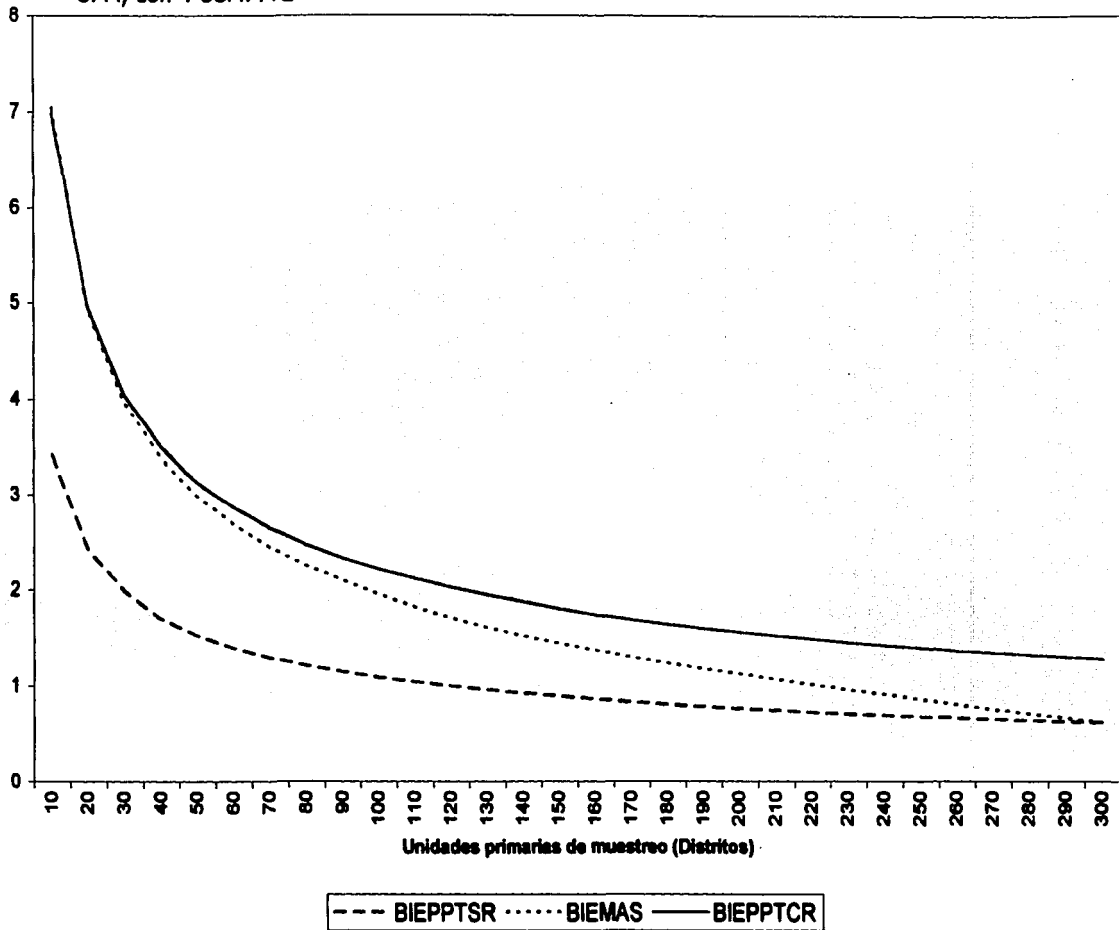
* Las coordenadas indican UPM y USM en el muestreo bietápico: (UPM, USM)

Figura 3.5. Error de estimación $\delta \times 10^2$ ($\delta = 1.96\sqrt{ECM(\hat{R})}$) para diseños bietápico varios tamaños de UPM, con 4 USM. Alianza por el Cambio



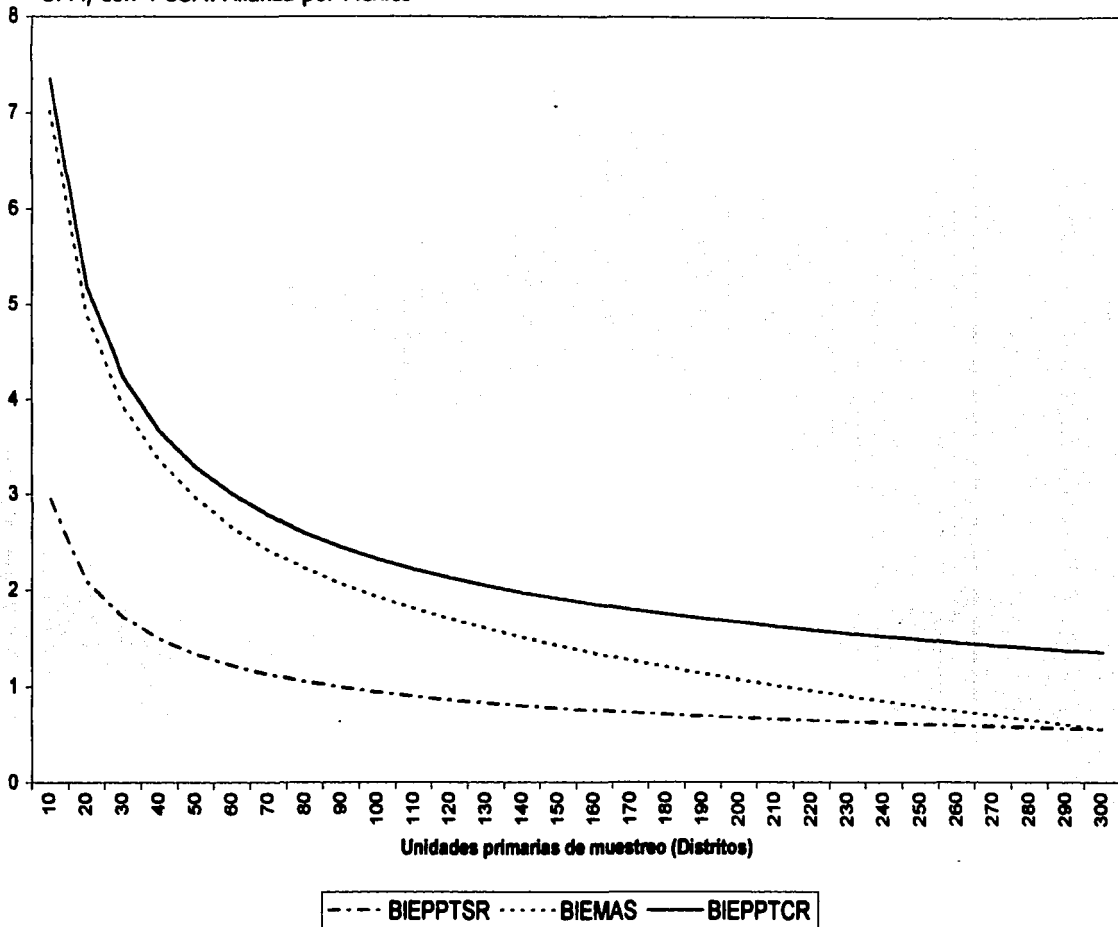
INSTITUTO VENEZOLANO DE ESTADÍSTICA
 IVE
 CAROLINA GUZMÁN
 MÓDULO 3

Figura 3.6. Error de estimación $\delta \times 10^2$ ($\delta = 1.96 \sqrt{ECM(\hat{R})}$) para diseños bietápicos varios tamaños de UPM, con 4 USM. PRI



TESIS CON
 FALTA DE CALIDAD

Figura 3.7. Error de estimación $\delta \times 10^2$ ($\delta = 1.96 \sqrt{ECM(\hat{R})}$) para diseños bietápicos varios tamaños de UPM, con 4 USM. Alianza por México



TESIS CON
 FALLA DE ORIGEN

3.3 Muestreo bietápico

Al aplicar el muestreo bietápico con selección aleatoria simple en ambas etapas (BIEMAS) con los distritos electorales como unidades primarias, en particular cuando se consideran los 300 distritos, se obtiene el error cuadrático medio menor para los diferentes partidos políticos después del E300S. Lo anterior motiva que en este apartado se analicen algunos resultados al evaluar la ecuación 1.8 del ECM del muestreo bietápico además de observar por separado los dos sumandos del ECM donde el primero corresponde a la variabilidad de las Unidades Primarias de Muestreo (UPM), es decir a la variabilidad entre distritos electorales y el segundo a la variabilidad de las secciones electorales, es decir las Unidades Secundarias de Muestreo (USM).

Con estos consideraciones se realizó en primer lugar la gráfica que se presenta en la figura 3.8 del error de estimación δ con el total del ECM calculado para cada partido al considerar muestras de 40, 60, 100, 150, 200 y 300 distritos; en cada distrito el número de secciones en muestra varía de 2 a 20.

Se puede observar que el error de estimación, al igual que en todos los demás diseños, es más grande para la Alianza por el Cambio, mientras que para el PRI y la Alianza por México el valor es similar, diferenciándose sólo de acuerdo al número de USM en muestra. Para otros partidos el valor de δ resultó pequeño en comparación con lo calculado para los partidos principales.

Se confirma un decrecimiento del valor del error de estimación δ , por consiguiente del ECM, en relación con el aumento de UPM. De manera similar, fijando el número de UPM el valor de δ disminuye respecto al aumento de secciones electorales en muestra. En los puntos donde cambia el tamaño de muestra de UPM se observa que el error de estimación sube con respecto a los valores anteriores, sin embargo la tendencia decreciente se sigue manteniendo rápidamente; esto se debe a que el tamaño total de muestra es mayor en el punto anterior. Para ilustrar lo anterior obsérvese en la figura 3.8 el pico al inicio de 200 distritos que corresponde a 400 secciones electorales en muestra (200 UPM y 2

USM) mientras que el punto inmediato anterior corresponde a una muestra total de 3000 secciones electorales (150 UPM y 20 USM).

En la figura 3.9 se presenta la gráfica del error de estimación correspondiente al primer término de la ecuación 1.8 para el ECM, es decir la que considera la variabilidad de las UPM (distritos). Se puede observar que el comportamiento de esta parte del error de estimación es similar al total, donde para la Alianza por el Cambio se presentan valores mayores; por otra parte se observa que el error de estimación para la Alianza por México presentan valores ligeramente mayores que el PRI, lo cual nos indicaría un voto más homogéneo para éste partido en los distritos electorales.

La gráfica que contiene los valores del error de estimación debida a la variabilidad de las USM (secciones), segundo término de la ecuación 1.8, se puede observar en la figura 3.10. El error de estimación para la Alianza por el Cambio sigue el mismo patrón que en las dos gráficas anteriores en el sentido de tener valores más altos que los partidos restantes, sin embargo ahora la diferencia parece ser más pequeña. En oposición a lo que sucede con la variabilidad debida a las UPM, en el caso de las USM es mayor en el PRI respecto a la de la Alianza por México, lo cual nos indica una mayor heterogeneidad para el PRI al interior de los distritos.

Para analizar el comportamiento del ECM total y de los términos que corresponde a la variabilidad de las UPM y USM en cada partido se presentan las figuras 3.11, 3.12, 3.13 y 3.14 para Alianza por el Cambio, PRI, Alianza por México y otros partidos respectivamente. En estas gráficas se puede observar que el aporte principal a los valores del ECM en el muestreo bietápico se debe a la variabilidad entre las UPM, es decir a los distritos electorales.

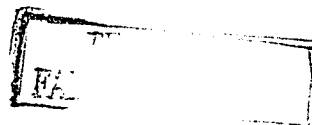


Figura 3.8. Error de estimación $\delta \times 10^2$, $\delta = 1.96\sqrt{ECM(\hat{R})}$, en el muestreo bietápico (BIEMAS), con Distritos como UPM y Secciones como USM, para los diferentes partidos.

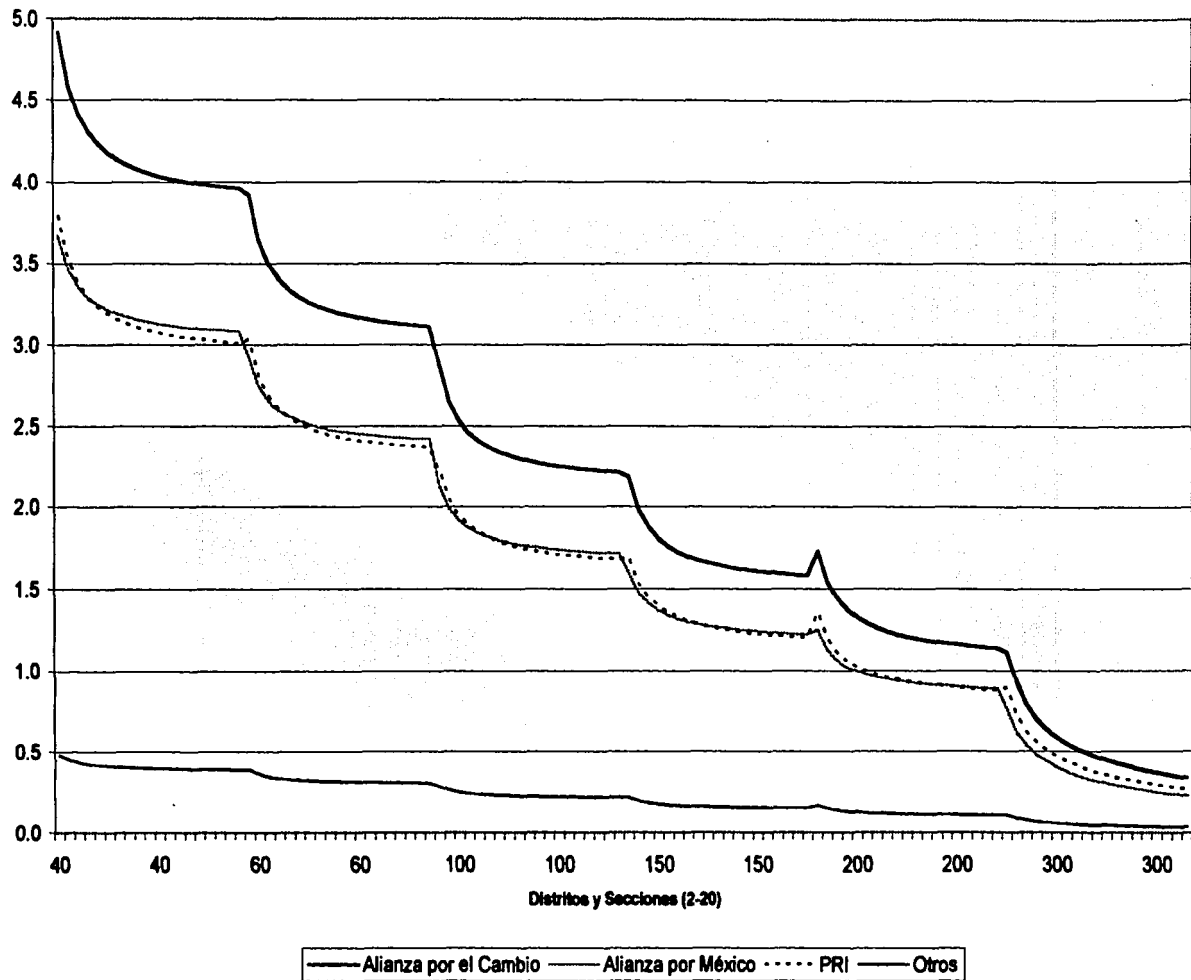
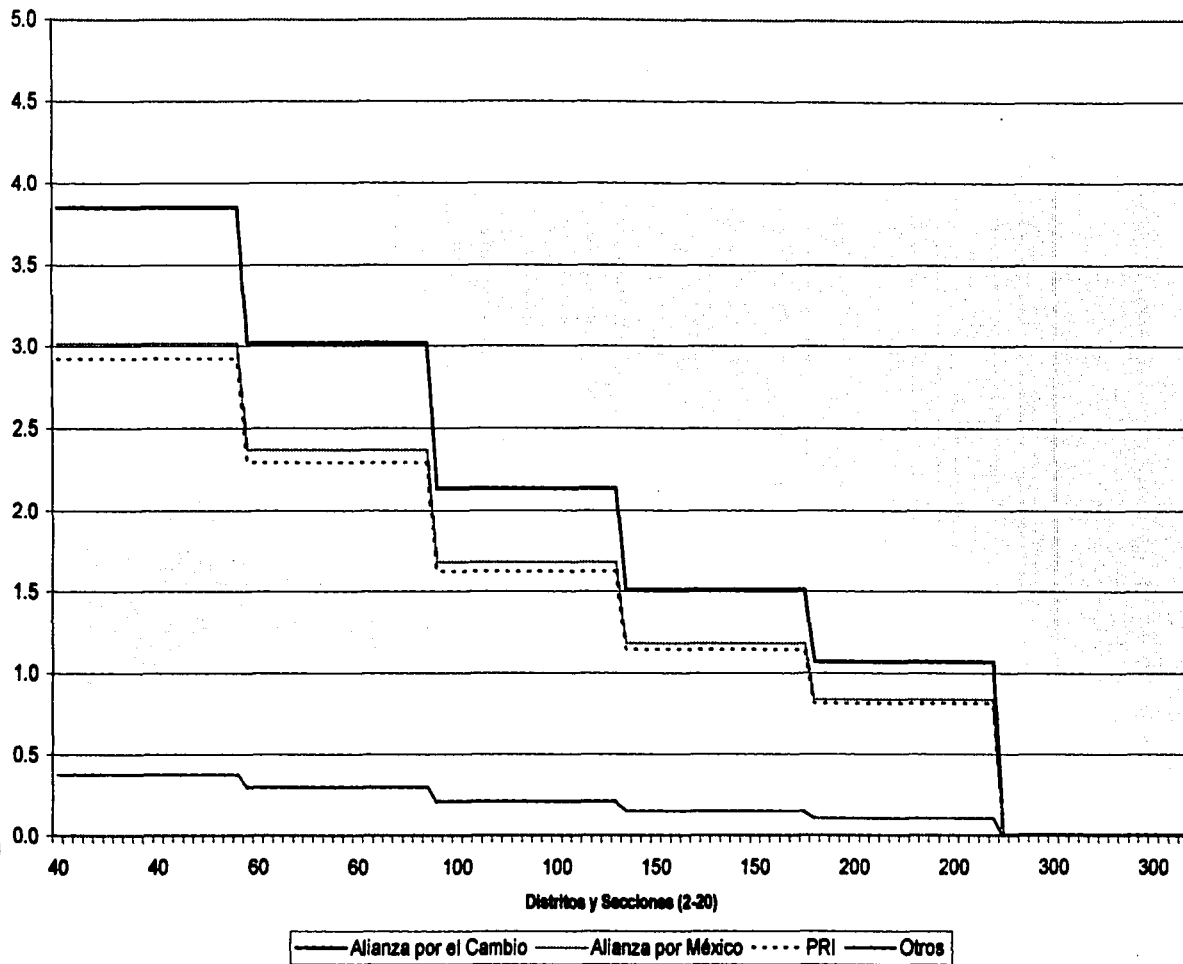


Figura 3.9. Error de estimación correspondiente a Unidades Primarias de Muestreo (Distritos), primer término de la expresión 1.8, $\delta \times 10^2$, $\delta = 1.96\sqrt{ECM(\hat{R})}$, en el muestreo bietápico (BIEMAS), para los diferentes partidos.



TESIS CON
 FALTA DE CATEGORIAS
 NO CISEL

Figura 3.10. Error de estimación correspondiente a Unidades Secundarias de Muestreo (Secciones), segundo término de la expresión 1.8, $\delta \times 10^2$, $\delta = 1.96 \sqrt{ECM(\hat{R})}$, en el muestreo bietápico (BIEMAS) para los diferentes partidos.

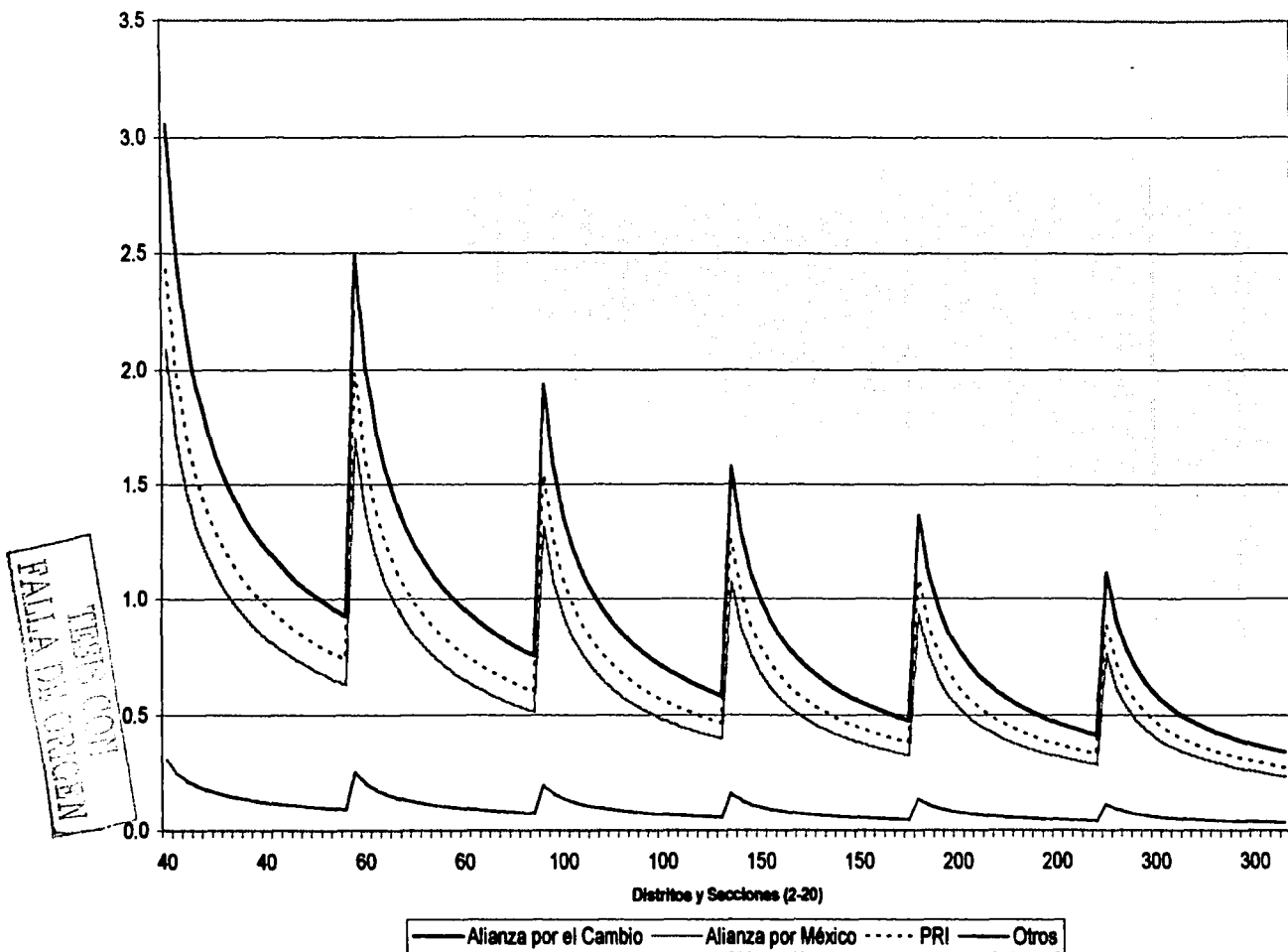
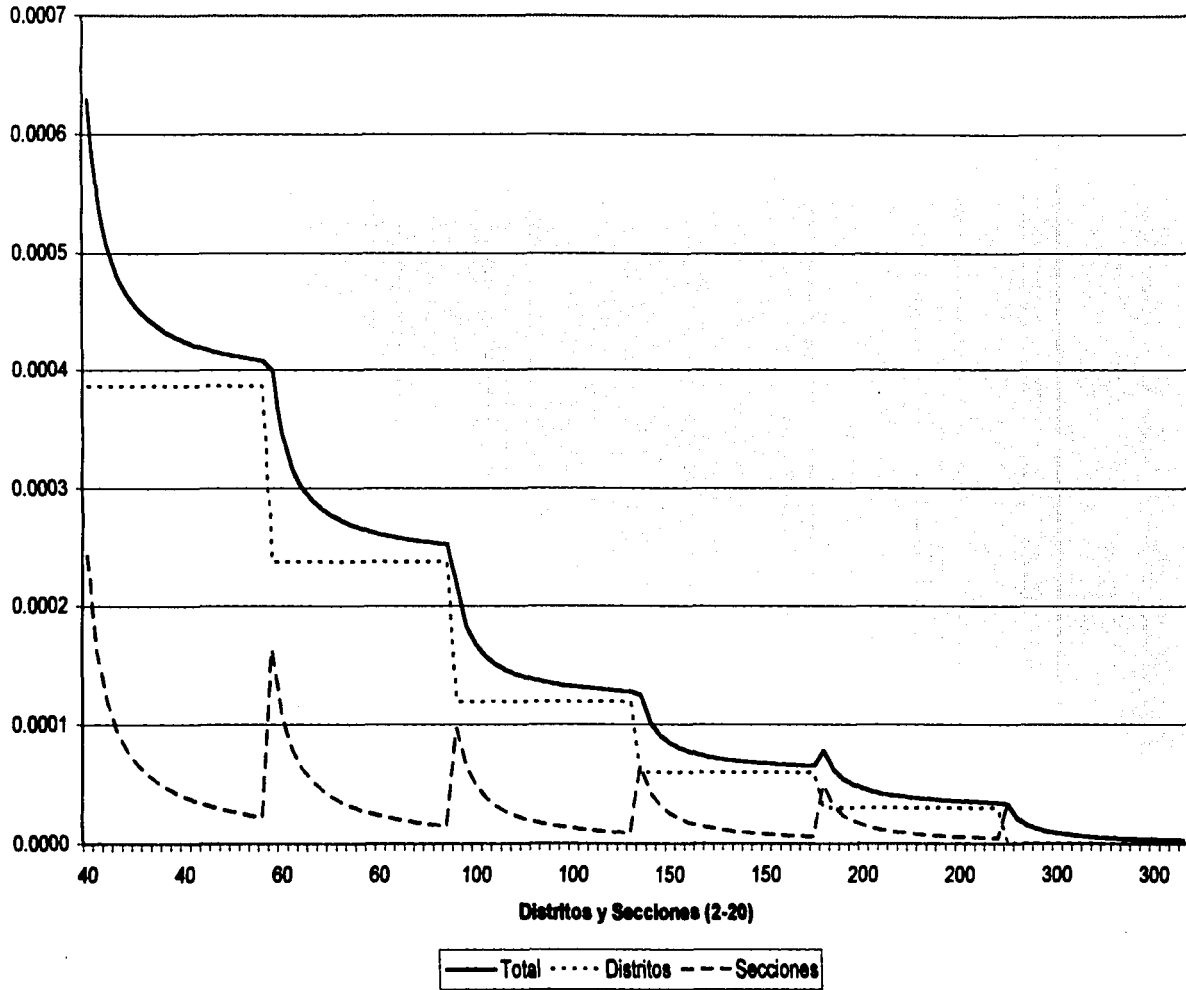
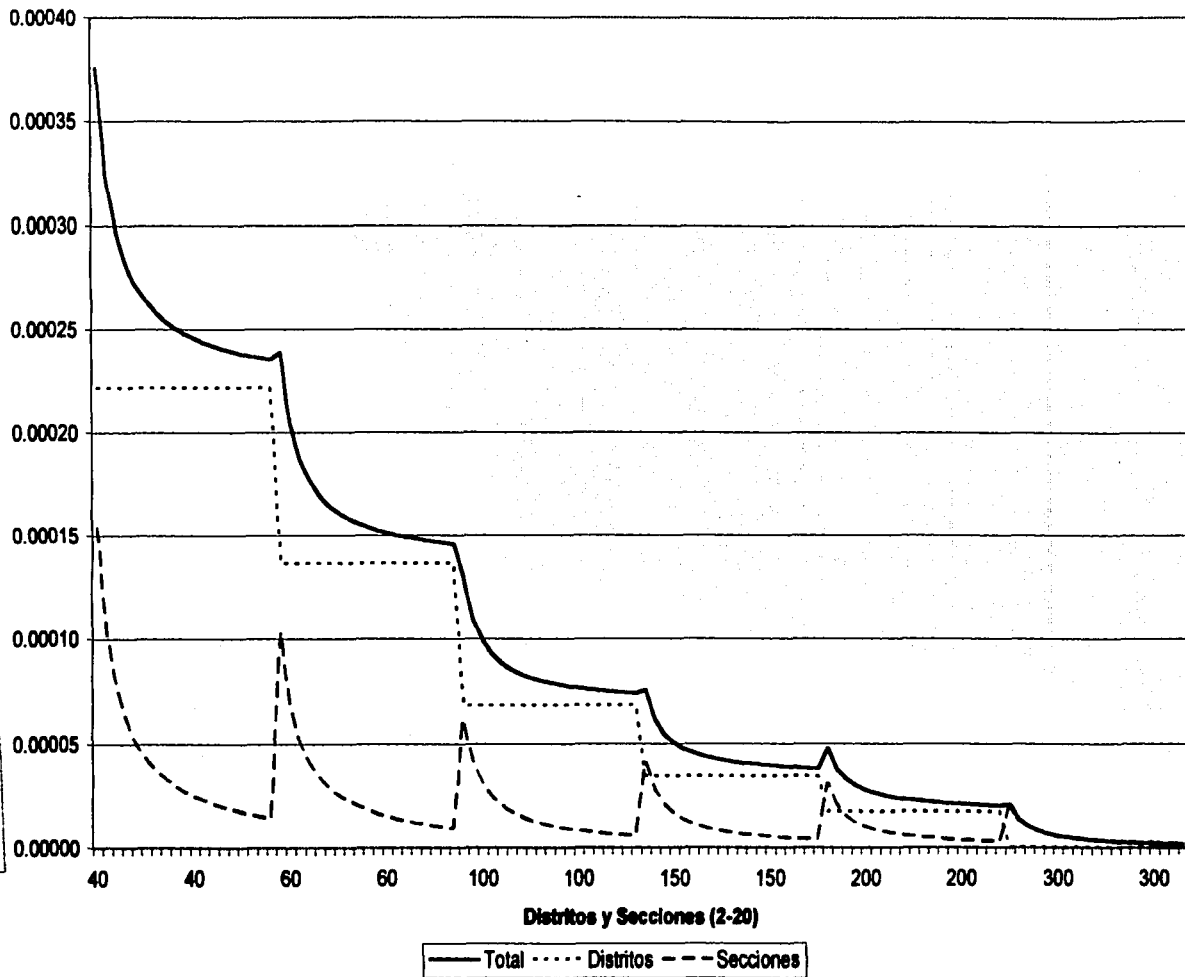


Figura 3.11. Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM(Distritos) y de USM(Secciones). Alianza por el Cambio.



TIRAR CON
 FALTA DE REGISTRO
 AGO 2011

Figura 3.12. Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM(Distritos) y de USM(Secciones). PRI.



TITULO COM
 PATENTE DE INGEN

Figura 3.13. Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM(Distritos) y de USM(Secciones). Alianza por México.

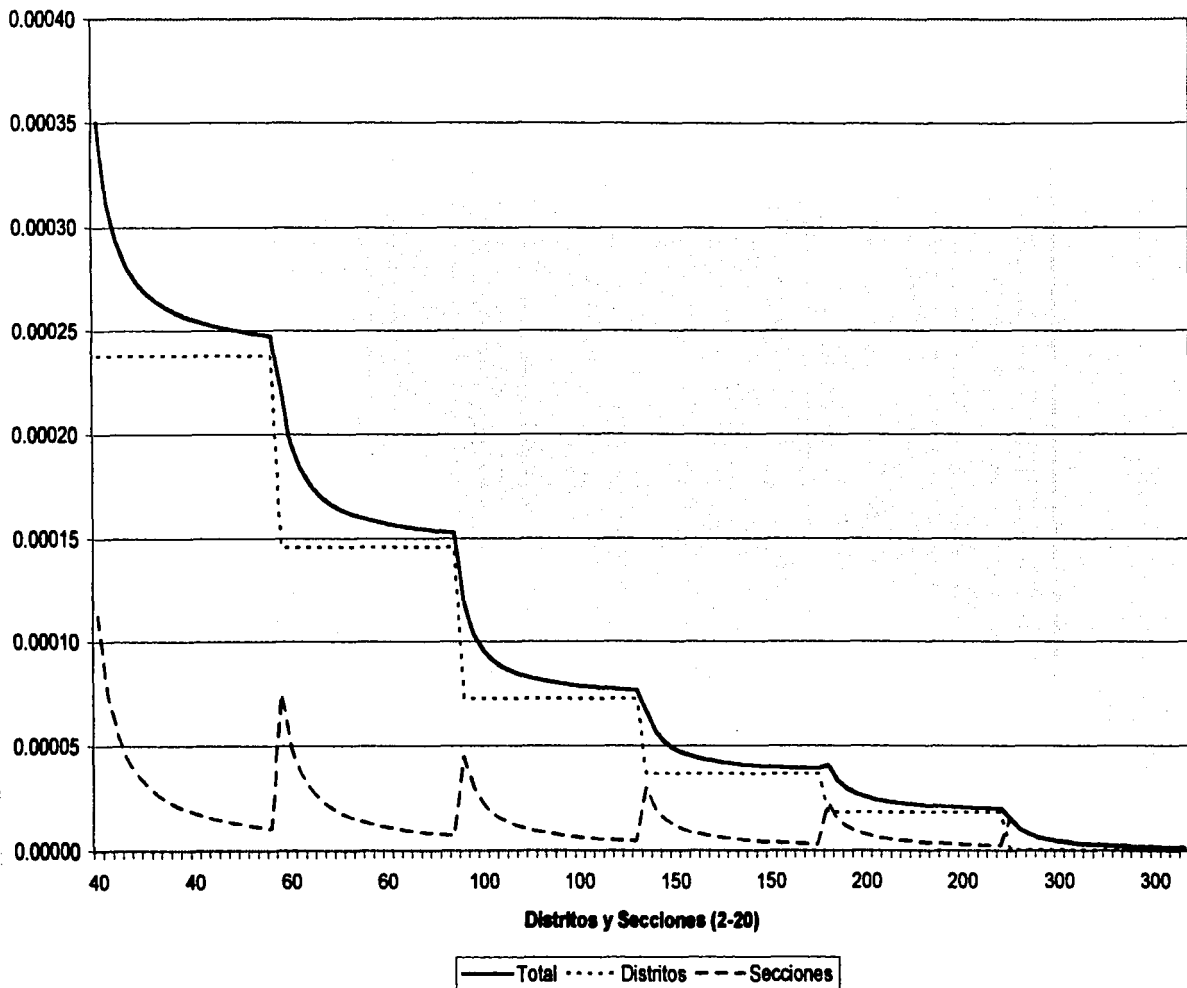
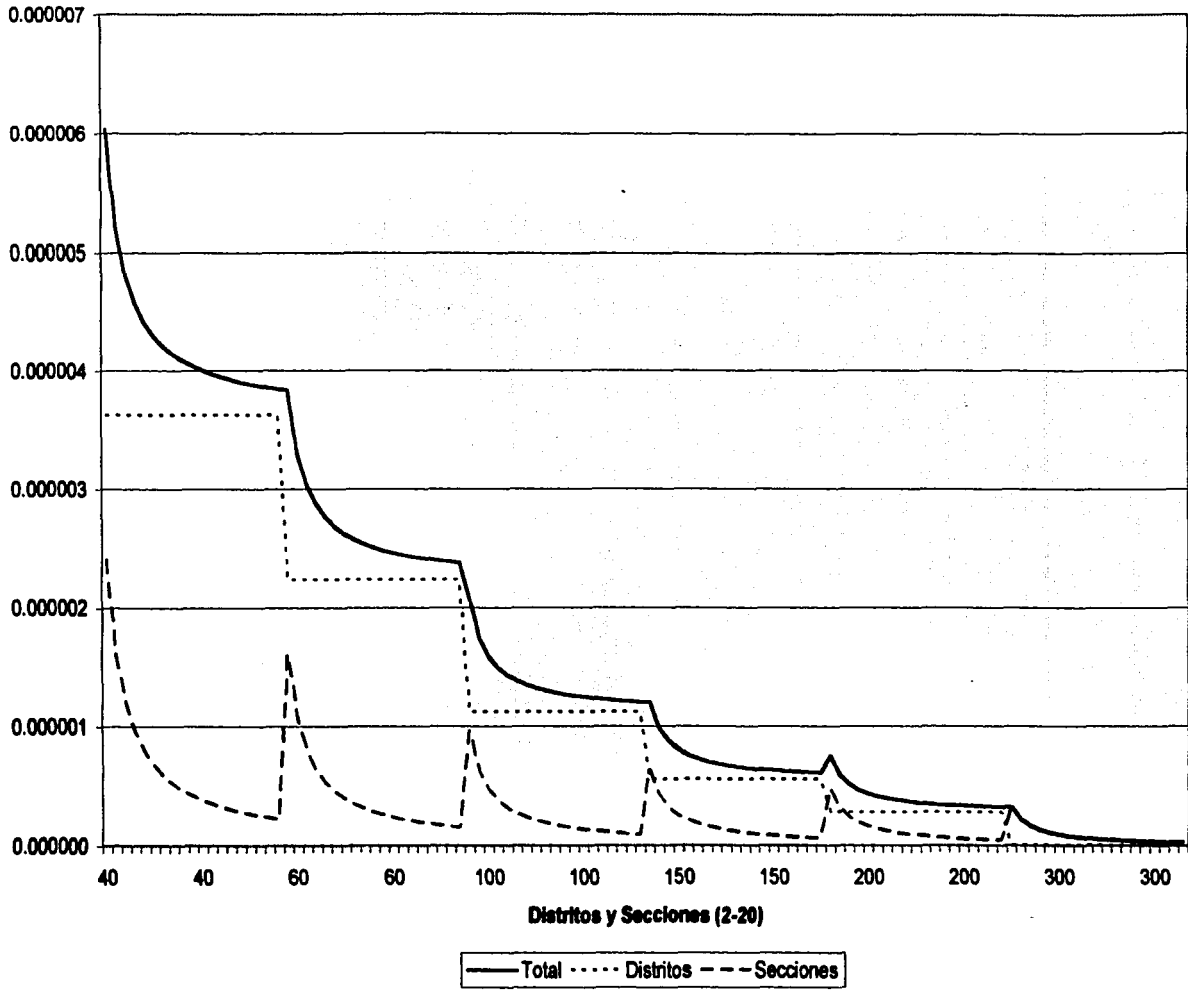


Figura 3.14. Error cuadrático medio en el diseño bietápico (BIEMAS), aportación de UPM(Distritos) y de USM(Secciones). Otros partidos.



755001
 FALLA DE PLAN

Figura 3.15. Error Cuadrático Medio en el muestreo bietápico (BIEMAS) para unidades primarias y secundarias de muestreo (2 y 3 USM en cada UPM). Alianza por el Cambio

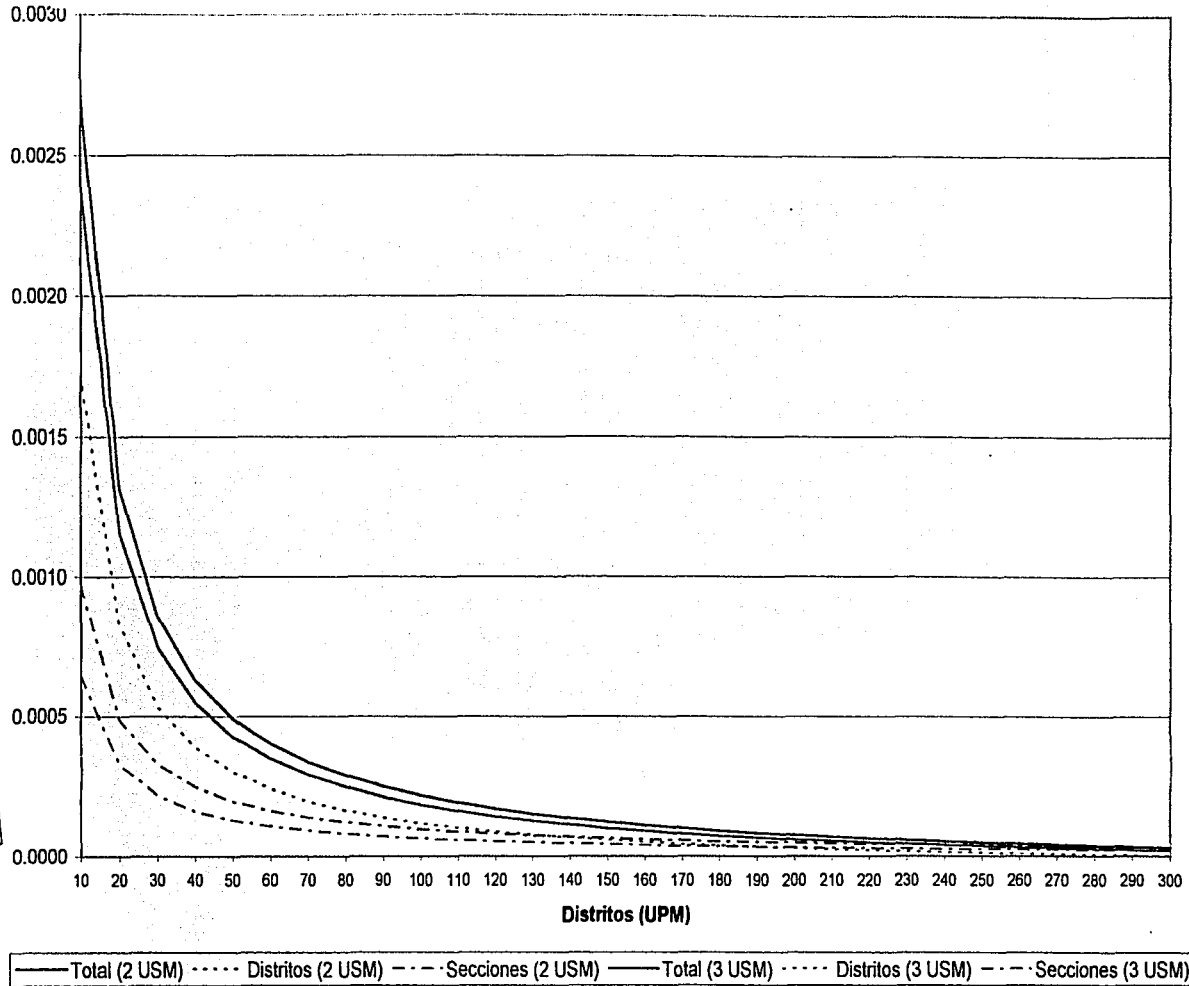


Figura 3.16. Error Cuadrático Medio en el muestreo bietápico (BIEMAS) para unidades primarias y secundarias de muestreo (2 y 3 USM en cada UPM), PRI

TESIS CON
 FALLA DE ORIGEN

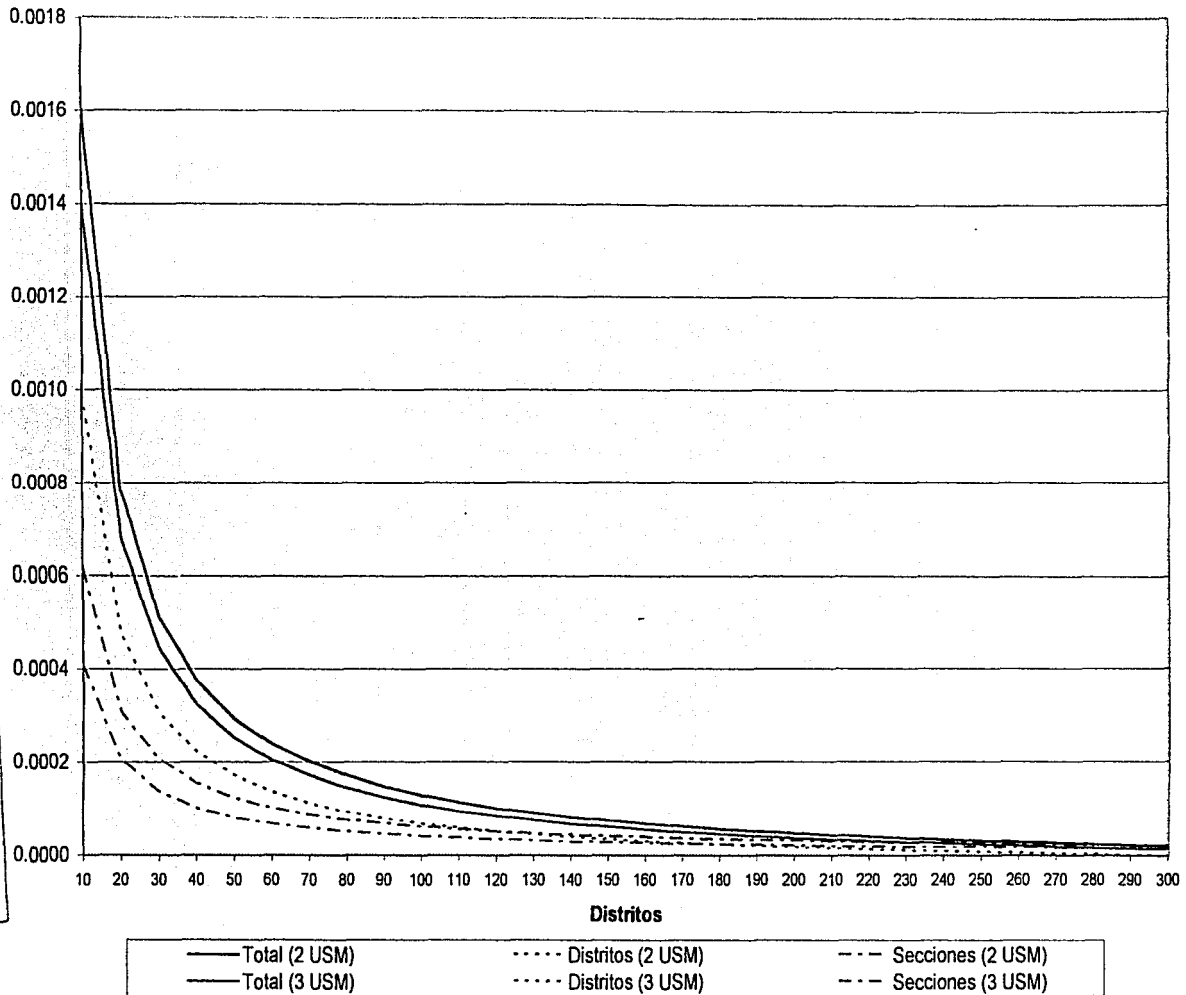
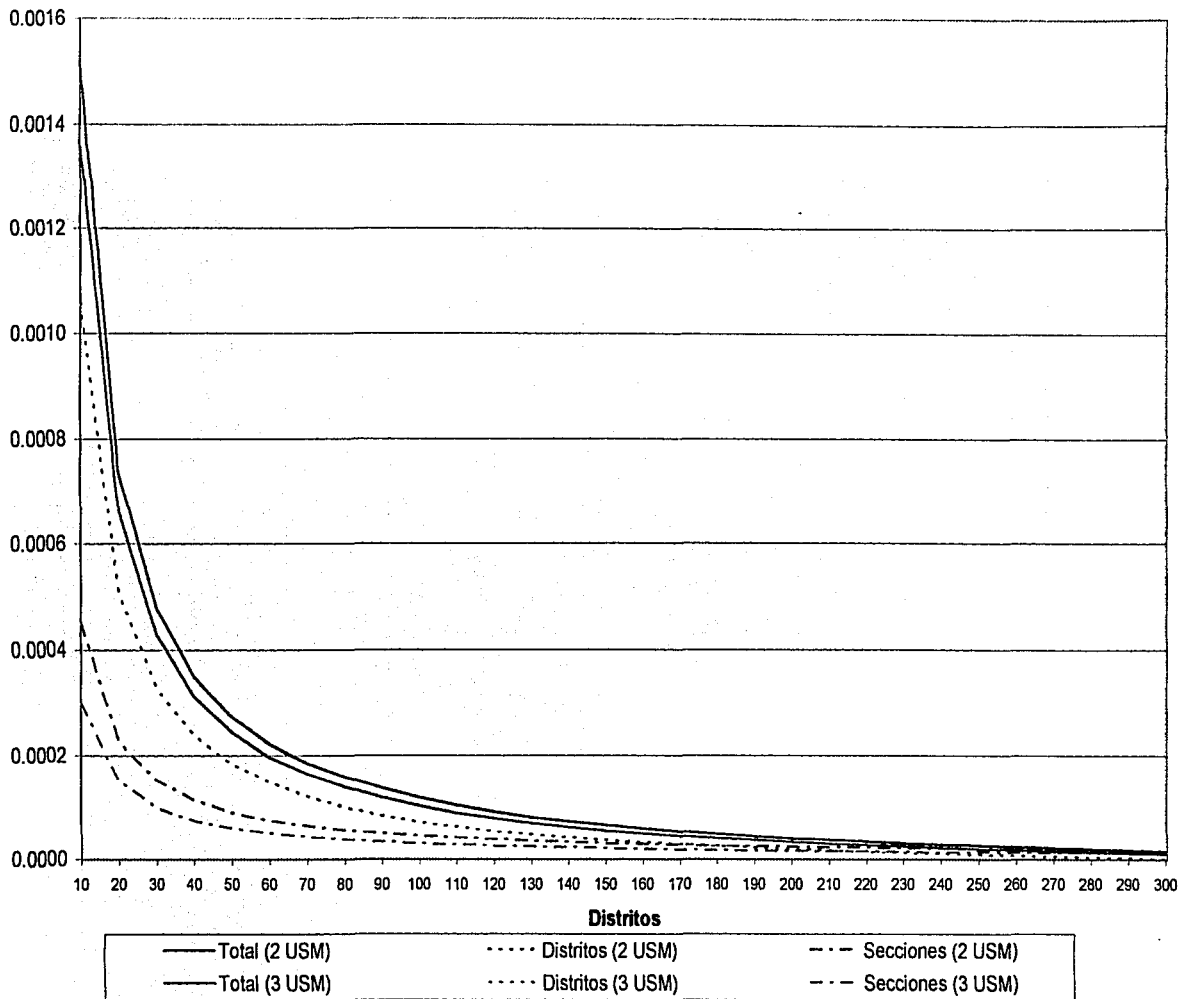


Figura 3.17. Error Cuadrático Medio en el muestreo bietápico (BIEMAS) para unidades primarias y secundarias de muestreo (2 y 3 USM en cada UPM). Alianza por México



INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS

Por otra parte y con el interés de averiguar cuál es el tamaño de muestra conveniente de UPM se evaluó la ecuación 1.8 para el error cuadrático medio variando el número de UPM entre 10 y 300 con muestras de 2 y 3 USM en cada distrito. En las figuras 3.15, 3.16 y 3.17 se presentan las gráficas para Alianza por el Cambio, PRI y Alianza por México respectivamente. Las líneas para el caso de distritos (UPM) en estas gráficas es similar en el caso de las 2 y 3 USM, por lo cual sólo se observan cinco líneas.

Atendiendo a lo observado en estas gráficas vemos que el decrecimiento del error cuadrático medio es insignificante después de 150 distritos. Sin embargo tomar una muestra de UPM de al menos este tamaño podría ser muy costoso en términos de trabajo de campo. Otro aspecto observado en las gráficas es la disminución drástica del valor del ECM entre 10 y 90 distritos aproximadamente. Por lo tanto, sería recomendable, si la decisión es considerar el diseño bietápico, tomar 90 distritos o más UPM y como considerar 2 o 3 USM prácticamente ya no afecta al ECM, entonces se recomiendan 2.

En la figura 3.18 se presenta el error de estimación, $\delta = 1.96\sqrt{ECM(\hat{R})}$, variando el número de USM de 2 a 5 y el número de distritos de 10 a 300. Comparando las curvas correspondientes a Alianza para el Cambio se observa que al aumentar una sección de muestra en cada distrito la diferencia se hace cada vez más estrecha. Es decir, la ganancia en precisión al pasar de 4 a 5 secciones es menor que la de pasar de 2 a 3. La misma afirmación puede realizarse para los restantes partidos, lo cual indicaría que en la práctica se debe considerar un número no muy grande de USM, recomendándose 2.

3.4 Error de estimación centrado en el parámetro para los diferentes diseños muestrales

El error cuadrático medio puede ser utilizado para calcular intervalos centrados en el parámetro de interés, mediante la expresión $R \pm 1.96\sqrt{ECM(\hat{R})}$ y que llamaremos error de estimación centrado en el parámetro. Este intervalo no debe confundirse



con la estimación por intervalo de confianza del parámetro de interés que tiene la expresión $\hat{R} \pm 1.96\sqrt{ECM(\hat{R})}$ el cual se centra en el estimador del parámetro.

En la figura 3.19 se presenta la gráfica del error de estimación centrado en el parámetro para los diseños muestrales considerados comparando para cada partido y tomando diferentes tamaños de muestra. Como ya se había comentado en el apartado 3.2, la mejor estimación en términos de precisión se logra mediante el diseño estratificado por distritos realizando la estimación separada (E300S) y en contraste el muestreo aleatorio simple (MAS) proporciona errores de estimación mayores. El elegir un diseño u otro para un efecto práctico deberá considerar, además de los valores del ECM, criterios relacionados con el costo de su implementación en el campo.

En la figura 3.19 se puede observar que en ningún caso existen intersecciones entre los intervalos centrados en R calculados para los diferentes partidos, lo cual nos lleva a pensar que si se hubieran aplicado los diseños de muestreo en la elección presidencial de julio del 2000 las diferencias en las proporciones entre los partidos resultarían significativas y la puesta en marcha de cualquiera de los diseños hubiera llevado a determinar con cierta confiabilidad la tendencia presentada en cuanto a los resultados obtenidos en la elección.

Para visualizar con mayor amplitud el comportamiento del error de estimación centrado en el parámetro de interés para cada uno de los partidos en las figuras 3.20, 3.21 y 3.22 se presentan las gráficas correspondientes a Alianza por el Cambio, PRI y Alianza por México respectivamente.

Tomando en cuenta que en la práctica el diseño que podría ser más económico es el bietápico, se calculó el error de estimación con diferentes combinaciones de UPM (distritos) y USM (secciones) y se graficaron en la figura 3.23 para los diferentes partidos. Puede observarse que los intervalos para el PRI y Alianza por el Cambio se cruzan en el caso 40 distritos y 15 secciones ($n=600$), lo cual nos indica que en la práctica se deben considerar más distritos (UPM) en muestra. En general puede observarse que el incremento en la muestra de UPM

disminuye el error de estimación, lo cual implica una mejor precisión, sin embargo los costos en la práctica aumentarían, por lo cual se tendrá que buscar un equilibrio entre estos dos aspectos.

3.4 Sesgo en el muestreo aleatorio simple

En el capítulo 1 se mencionó que el estimador \hat{R} de R es sesgado. El sesgo de este estimador se presenta en la expresión 3.7. Se puede demostrar (Cochran 1980, que una cota para el sesgo relativo esta dada por el coeficiente de variación

de \bar{y} . Es decir: $\frac{\text{sesgo}(\hat{R})}{\sqrt{\text{var}(\hat{R})}} \leq CV(\bar{y})$. Un criterio establecido por este autor es que el

sesgo es despreciable si el coeficiente de variación es menor de 0.1. Al calcular y graficar estas dos expresiones para varios tamaños de muestra, ver figura 3.24, se puede concluir que en el caso de la Alianza por el Cambio para las estimaciones que se han realizado los sesgos son despreciables. Misma situación sucede con las estimaciones de los otros partidos.

Figura 3.18. Error de estimación $\delta \times 100$, $\delta = 1.96\sqrt{ECM(\hat{R})}$ en el muestreo bietápico (BIEMAS) para unidades primarias y secundarias de muestreo (2 a 5 USM en cada UPM).

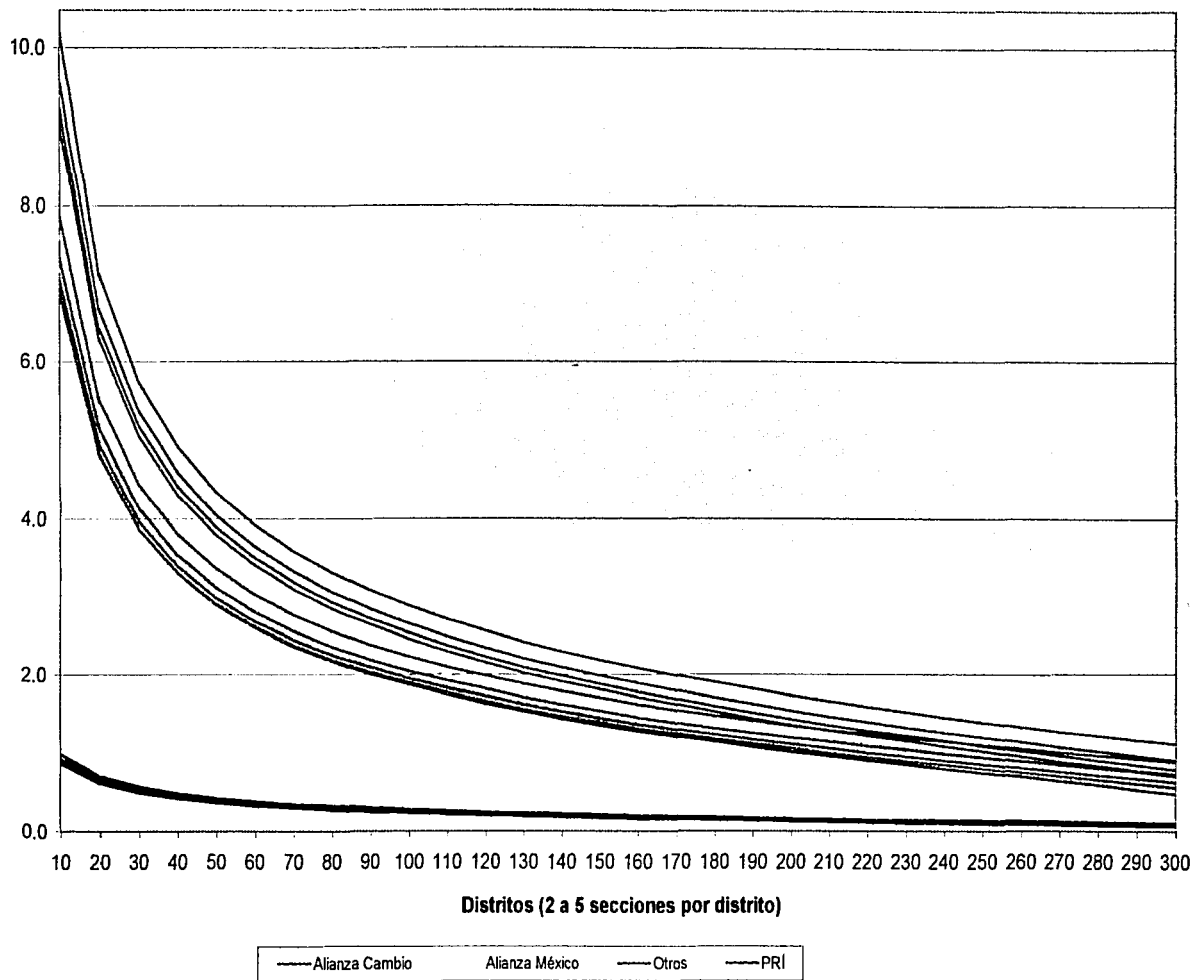
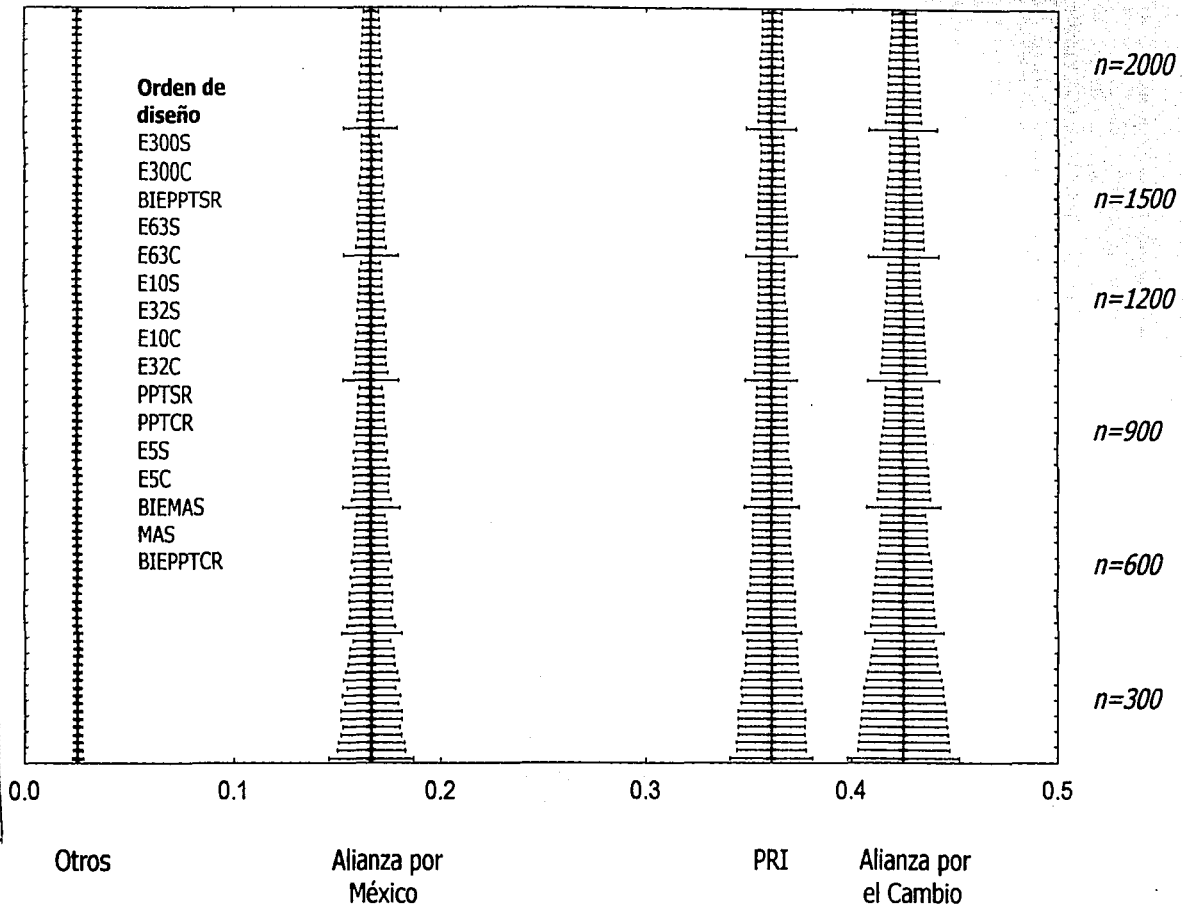


Figura 3.19. Error de estimación centrado en el parámetro de interés ($R \pm 1.96\sqrt{ECM(\hat{R})}$)* de cada partido, bajo diferentes diseños de muestreo.**

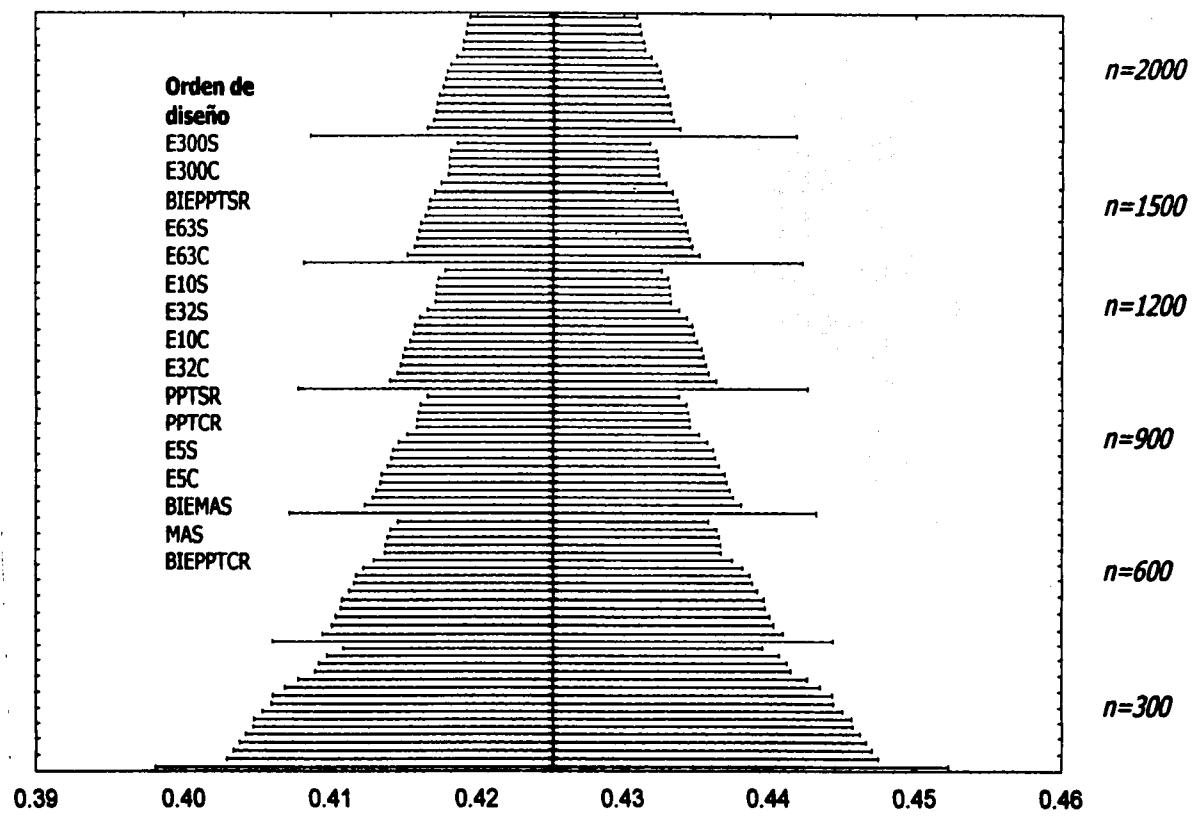


FALTA... TESIS CON...

* No confundir con intervalo de confianza para el parámetro de interés $\hat{R} \pm 1.96\sqrt{ECM(\hat{R})}$

** Para los diseños bietápicos, el número de distritos (UPM) muestreados, correspondiente a cada tamaño de muestra es: 150 (n=300); 300 (n=600), 300 (n=900); 300 (n=1200); 300 (n=1500); 300 (n=2000).

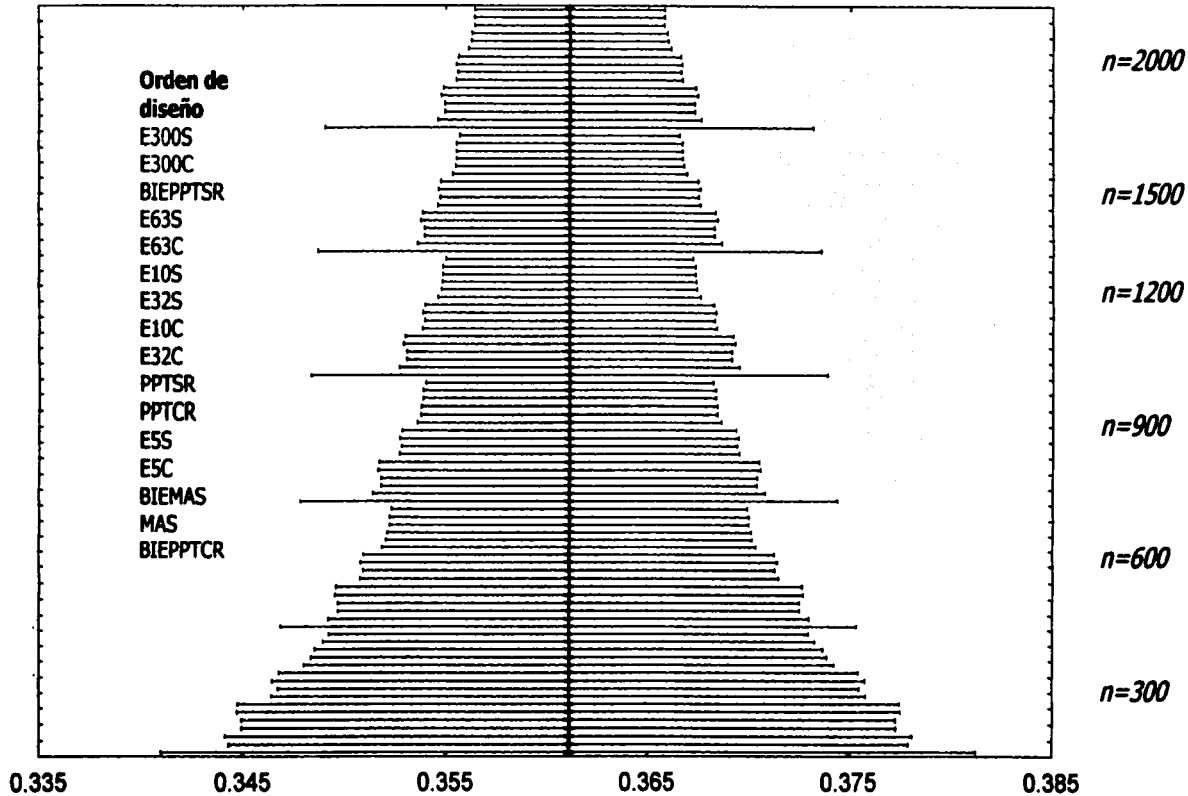
Figura 3.20. Error de estimación centrado en el parámetro de interés $(R \pm 1.96 \sqrt{ECM(\hat{R})})^*$ bajo diferentes diseños de muestreo** Alianza por el Cambio.



* No confundir con intervalo de confianza para el parámetro de interés $\hat{R} \pm 1.96 \sqrt{ECM(\hat{R})}$

** Para los diseños bietápicos, el número de distritos (UPM) muestreados, correspondiente a cada tamaño de muestra es: 150 (n=300); 300 (n=600), 300 (n=900); 300 (n=1200); 300 (n=1500); 300 (n=2000).

Figura 3.21 Error de estimación centrado en el parámetro de interés($R \pm 1.96\sqrt{ECM(\hat{R})}$) * bajo diferentes diseños muestrales.** PRI

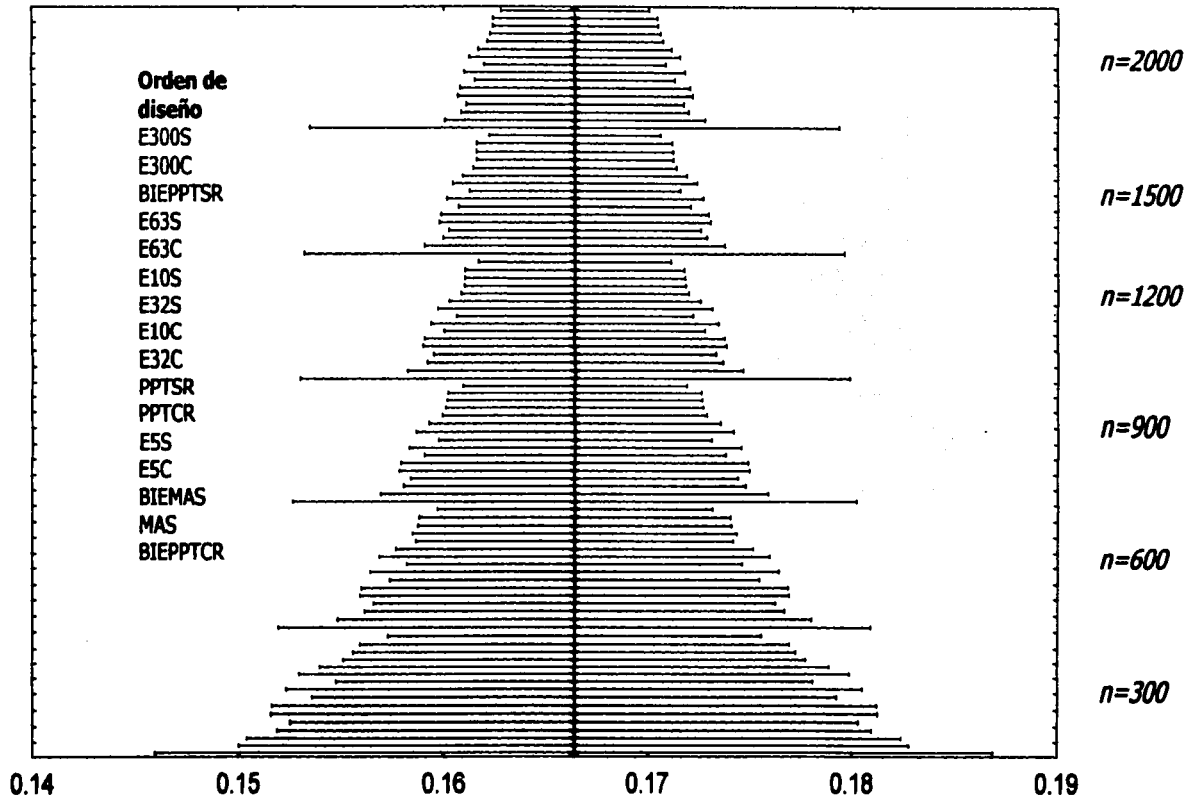


* No confundir con intervalo de confianza para el parámetro de interés $\hat{R} \pm 1.96\sqrt{ECM(\hat{R})}$

** Para los diseños bietápicos, el número de distritos (UPM) muestreados, correspondiente a cada tamaño de muestra es: 150 (n=300); 300 (n=600), 300 (n=900); 300 (n=1200); 300 (n=1500); 300 (n=2000).

TESIS CON FALTA DE PAGINA

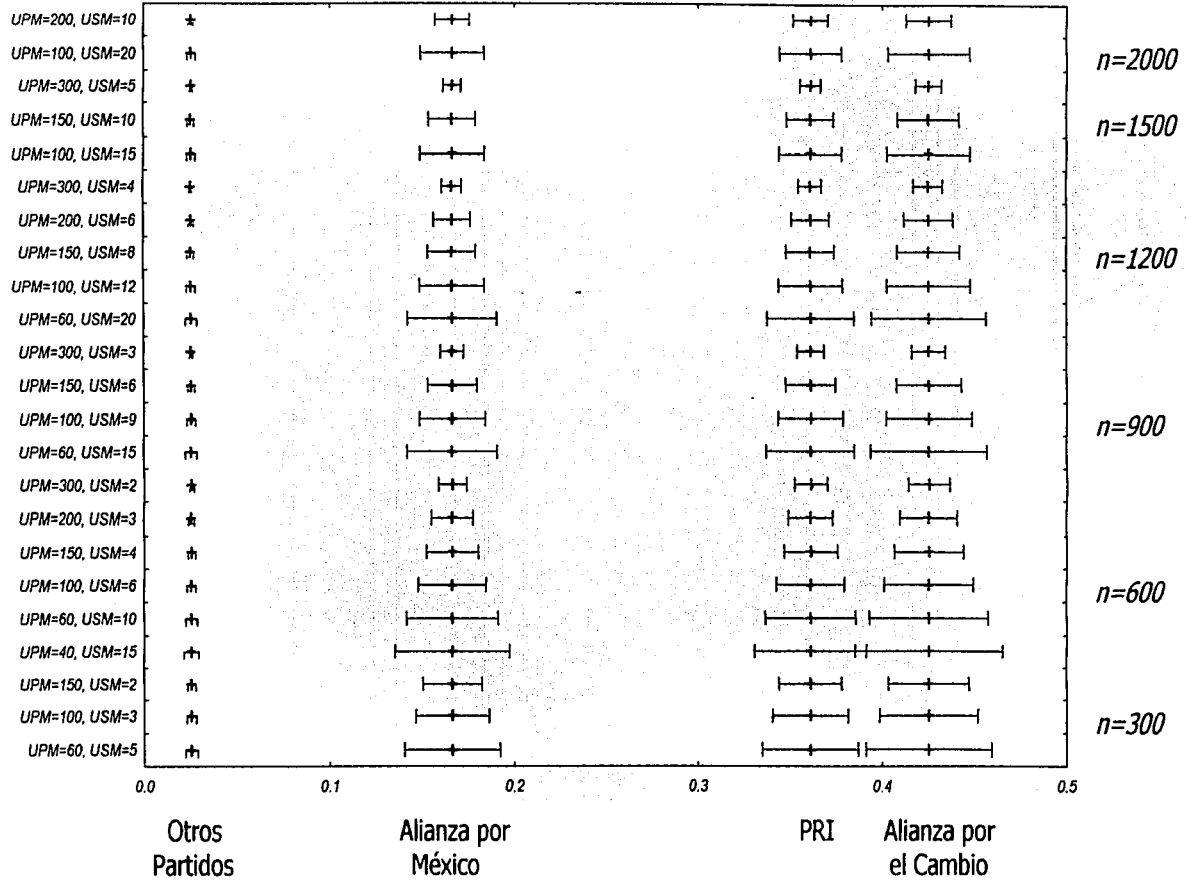
Figura 3.22. Error de estimación centrado en el parámetro de interés ($R \pm 1.96\sqrt{ECM(\hat{R})}$)* bajo diferentes diseños de muestreo. ** Alianza por México



* No confundir con intervalo de confianza para el parámetro de Interés $\hat{R} \pm 1.96\sqrt{ECM(\hat{R})}$

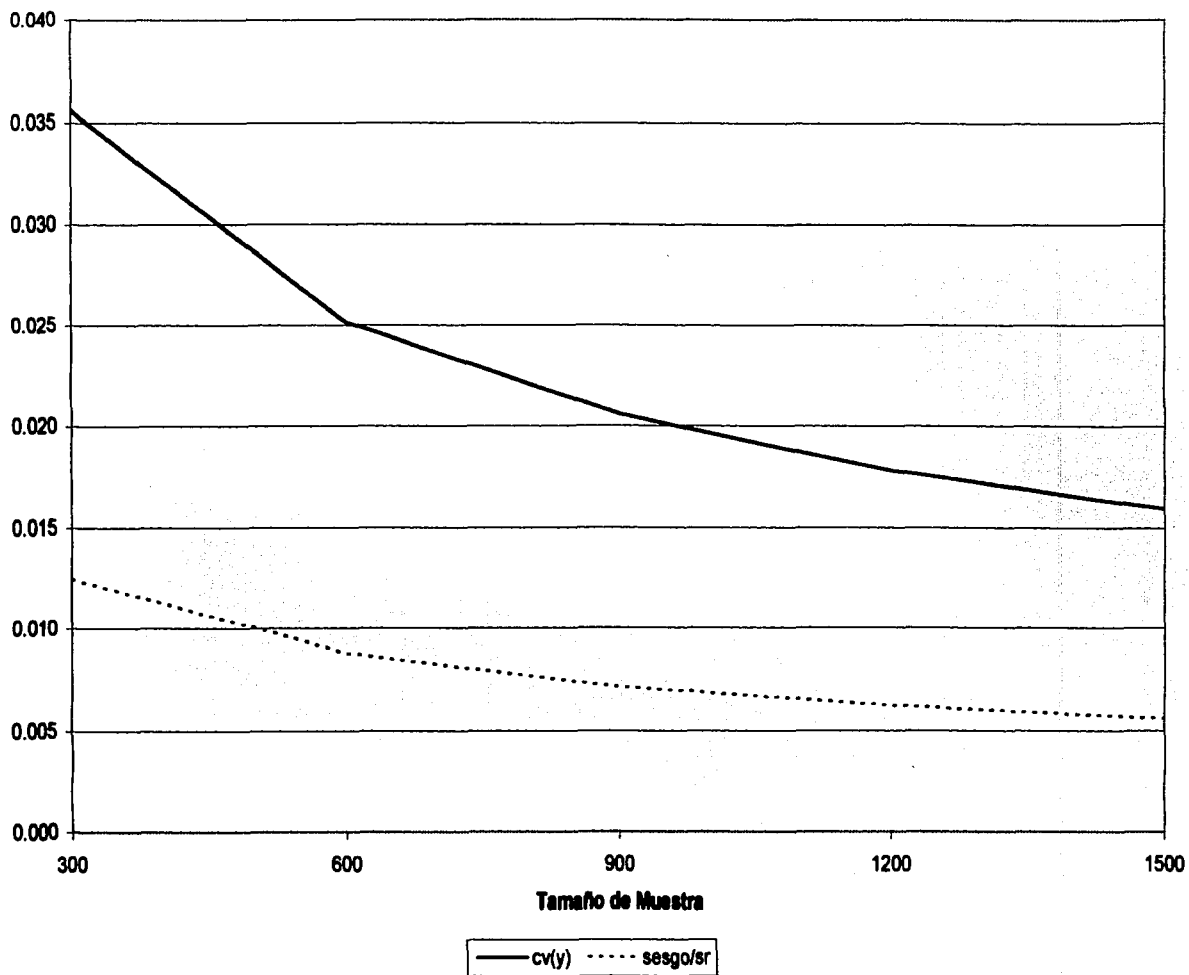
** Para los diseños bietápico, el número de distritos (UPM) muestreados, correspondiente a cada tamaño de muestra es: 150 (n=300); 300 (n=600); 300 (n=900); 300 (n=1200); 300 (n=1500); 300 (n=2000).

Figura 3.23. Error de estimación centrado en el parámetro de interés $(R \pm 1.96 \sqrt{ECM(\hat{R})})^*$ de cada partido, Muestreo Bietápico (BIEMAS), diferentes tamaños de Unidades Primarias (Distritos) y Secundarias (Secciones) de Muestreo



*No confundir con intervalo de confianza para el parámetro de interés $\hat{R} \pm 1.96 \sqrt{ECM(\hat{R})}$

Figura 3.24. Sesgo relativo y coeficiente de variación en el muestreo aleatorio simple. Alianza por el Cambio

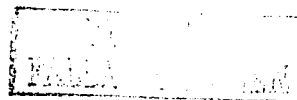


Capítulo 4. Conclusiones

Los diseños de muestreo trabajados en esta tesis consideran como marco de muestreo las 63445 secciones electorales a nivel federal, es decir son conglomerados de votantes. Estos conglomerados tienen características diferentes como es el hecho de pertenecer o no a áreas urbanas, la entidad donde se localizan, la circunscripción en donde se agrupo por parte del IFE, así como el distrito electoral al que pertenecen; todo esto propicia una enorme variabilidad en la composición del padrón electoral, lo que implica que el error cuadrático medio de los estimadores pueda ser muy grande. Lo anterior motiva que se propusieran varias maneras de estimar las proporciones de votación para los diferentes partidos. Se consideró que el total de votos emitidos en cada sección electoral es información adicional a los votos obtenidos por cada partido, por lo cual se podría utilizar el estimador de razón o el de regresión, para estimar la proporción de votos para cada partido.

En este trabajo se utilizó el estimador de razón bajo diferentes diseños de muestreo: aleatorio simple, con probabilidad proporcional al tamaño sin y con reemplazo, estratificado considerando dos estimadores: separado y combinado, definiendo los estratos por distritos electorales, por circunscripción, por entidad, por estado y tipo de sección (urbana y no urbana), por circunscripción y tipo de sección, bietápico considerando los distritos electorales como unidades primarias de muestreo y las secciones electorales como unidades secundarias; para este diseño se consideraron tres variantes: (1) muestreo aleatorio simple en las dos etapas, (2) con probabilidad proporcional al tamaño en la primera etapa con reemplazo y aleatorio simple en la segunda etapa y (3) con probabilidad proporcional al tamaño en la primera etapa sin reemplazo y aleatorio simple en la segunda etapa. En todos los diseños de muestreo se exploraron tamaños de muestra 300, 600, 900, 1200, 1500 y 2000.

La evaluación de los diferentes diseños se hizo a partir del cálculo del error cuadrático medio y la eficiencia relativa de los diferentes diseños con respecto al



muestreo aleatorio simple, $\frac{ECM(\hat{R}_{diseño})}{ECM(\hat{R}_{MAS})}$ con $n=1200$, obteniendo las siguientes

conclusiones:

- i) Los diseños que se basan en el muestreo estratificado resultaron ser más eficientes que el muestreo aleatorio simple, por ejemplo para la Alianza por el Cambio la eficiencia relativa del estratificado por circunscripción es de 1.14, mientras que para la estratificación por distritos (E300S) es de 2.27. Estos resultados confirman lo expresado teóricamente por algunos autores como Cochran (1993, pág. 136).
- ii) Los diseños de primera etapa que se basan en el muestreo con probabilidad proporcional al tamaño resultaron ser ligeramente mejores que el aleatorio simple, la eficiencia relativa es de 1.18 con reemplazo y de 1.21 sin reemplazo para Alianza por el Cambio. En algunos casos los diseños PPT son más eficientes que el estratificado con 5 circunscripciones como puede verse con lo comentado en el punto anterior.
- iii) Entre los diseños que se basan en el muestreo estratificado se esperaba que al aumentar el número de estratos en la población disminuyera el error cuadrático medio. Sin embargo no siempre se observó este patrón, por ejemplo en el caso de Alianza por el Cambio al considerar 10 estratos (5 circunscripción por 2 tipos de sección) se obtuvo 1.32 de eficiencia relativa mientras que al considerar 32 estratos (31 estados y el D.F.) se obtuvo 1.29, lo cual nos indica un valor del error cuadrático medio mayor con 32 estratos con relación a 10 estratos. Esto podría deberse a que la estratificación por distrito es muy diferente a la de estado y tipo y no tiene que ver tanto con el tamaño de los estratos.
- iv) En el muestreo estratificado la estimación basada en el estimador separado produce en todos los casos considerados mejores estimaciones que el estimador combinado, por ejemplo en la estratificación por distritos con el estimador separado la eficiencia relativa para el PRI es de 1.90 mientras que con el estimador combinado es de 1.79.

- v) El mejor diseño de muestreo, porque obtuvo un ECM más pequeño, resulto ser el que contempló el mayor número de estratos (300 distritos), considerando el estimador separado con una eficiencia relativa de 2.27 para Alianza por el Cambio, 1.90 para el PRI, 3.04 para Alianza por México y 2.15 para otros partidos.
- vi) La estratificación por estado y tipo de sección (63 estratos) produce errores ligeramente mayores, eficiencia relativa de 1.92, que el realizado por distritos, eficiencia de 2.27, por lo cual debería ser un diseño que en la práctica podría ser viable debido a que en términos operativos pudiera ser menos costoso.
- vii) El diseño basado en el muestreo bietápico con ambas etapas tomadas con muestreo aleatorio simple, con los distritos como UPM, produce errores similares al estratificado con 300 distritos, eficiencia relativa en el caso de Alianza por el Cambio de 2.01 y 2.27 respectivamente, por lo cual en la práctica debería ser considerado ya que ayudaría a reducir costos, siempre y cuando se considere una muestra de al menos 100 distritos. Lo mismo podría decirse del muestreo bietápico PPT sin reemplazo en la primera etapa (BIEPPTSR) donde la eficiencia relativa es de 1.96.
- viii) El diseño menos eficiente es el muestreo bietápico cuando las UPM se seleccionan con probabilidad proporcional al tamaño con reemplazo (BIEPPTCR) con una eficiencia relativa de 0.41 para Alianza por el Cambio y 0.43 para el PRI.
- ix) Los resultados de las elecciones presidenciales del año 2000 no fueron tan competidas como se esperaba, y de acuerdo a los diseños muestrales analizados un tamaño entre $n=600$ y $n=900$ secciones electorales hubiera sido suficiente para lograr inferir las proporciones de las votaciones con una buena confiabilidad. Ahora bien, si las elecciones se vislumbraran más competitivas quizás un tamaño de muestra entre 1200 y 1500 sea lo más apropiado.
- x) Con los resultados numéricos obtenidos se podría recomendar considerar un diseño estratificado por distrito o bietápico con los distritos como unidades de primera etapa de muestreo con selección PPT sin reemplazo, o un tamaño de

muestra variando de 900 a 1200 secciones, sin embargo no es suficiente dar argumentos sólo en términos estadísticos, ya que en la práctica puede ser muy costoso y de difícil control, además de verse influenciado por aspectos que tienen que ver con la composición de la población (rural, urbana) y con la competencia de la elección. Por lo tanto la respuesta sobre cuál diseño usar en el caso de una encuesta electoral presidencial a nivel nacional no es única, depende de factores estadísticos, de costos monetarios, de recursos humanos y de aspectos políticos del momento.

En esta tesis se ha evaluado el error cuadrático medio del estimador de razón para cada partido bajo distintos diseños, otra alternativa que se podría contemplar es la evaluación del error cuadrático medio de la diferencia de razones para los diferentes partidos o bien de la razón de razones que se explorará en trabajos posteriores.

También se tiene contemplado se para un trabajo posterior simular la obtención de una muestra en los diferentes diseños de muestreo y calcular el estimador de razón y el estimador del error cuadrático medio.

Actualmente se esta trabajando en la evaluación del error cuadrático medio en diseños basados en el muestreo sistemático.

TESIS CON
FALLA DE ORIGEN

ESTA TESIS NO SALE
DE LA BIBLIOTECA


```

700      format(i4, ' MAS ', 7x, 'Alianza Cambio', 1x, f7.4, 5x, f14.10)
      call mas(pri, total, n, ppri, vpri)
      write(7, 800) n, ppri, vpri
800      format(i4, ' MAS ', 7x, 'PRI', 12x, f7.4, 5x, f14.10)
      call mas(prd, total, n, pa_mex, va_mex)
      write(7, 900) n, pa_mex, va_mex
900      format(i4, ' MAS ', 7x, 'Alianza Mexico', 1x, f7.4, 5x, f14.10)

      call mas(otros, total, n, potros, votros)
      write(7, 950) n, potros, votros
950      format(i4, ' MAS ', 7x, 'Otros', 10x, f7.4, 5x, f14.10)

      ! Muestreo PPT Con Reemplazo
      call pptcr(pan, total, pi, n, pa_cam, va_cam)
      write(7, 1000) n, pa_cam, va_cam
1000     format(i4, ' PPTCR ', 4x, 'Alianza Cambio', 1x, f7.4, 5x, f14.10)
      call pptcr(pri, total, pi, n, ppri, vpri)
      write(7, 1100) n, ppri, vpri
1100     format(i4, ' PPTCR ', 6x, 'PRI', 11x, f7.4, 5x, f14.10)
      call pptcr(prd, total, pi, n, pa_mex, va_mex)
      write(7, 1200) n, pa_mex, va_mex
1200     format(i4, ' PPTCR ', 5x, 'Alianza México', 1x, f7.4, 5x, f14.10)

      call pptcr(otros, total, pi, n, potros, votros)
      write(7, 1250) n, potros, votros
1250     format(i4, ' PPTCR ', 4x, 'Otros ', 9x, f7.4, 5x, f14.10)
      ! Muestreo PPT Sin Reemplazo
      call pptsr(pan, total, pi, n, pa_cam, va_cam)
      write(7, 1001) n, pa_cam, va_cam
1001     format(i4, ' PPTSR ', 4x, 'Alianza Cambio', 1x, f7.4, 5x, f14.10)
      call pptsr(pri, total, pi, n, ppri, vpri)
      write(7, 1101) n, ppri, vpri
1101     format(i4, ' PPTSR ', 6x, 'PRI', 11x, f7.4, 5x, f14.10)
      call pptsr(prd, total, pi, n, pa_mex, va_mex)
      write(7, 1201) n, pa_mex, va_mex
1201     format(i4, ' PPTSR ', 5x, 'Alianza México', 1x, f7.4, 5x, f14.10)

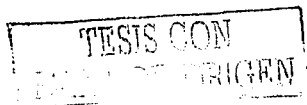
      call pptsr(otros, total, pi, n, potros, votros)
      write(7, 1251) n, potros, votros
1251     format(i4, ' PPTSR ', 4x, 'Otros ', 9x, f7.4, 5x, f14.10)
enddo
close(7)
end

```

```

subroutine mas(partido, total, n, r, var)
implicit none
integer::i, n
integer, parameter::double=selected_real_kind(p=13)
real(kind=double)::r, var, votos, tot
real::suma
integer, parameter::pobl=63445
real(kind=double), dimension(pobl)::partido, total
votos=0.
tot=0.
do i=1, pobl
votos=votos+partido(i)

```



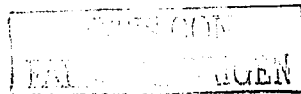
```

    tot=tot+total(i)
  end do
  r=votos/tot
  suma=0
  do i=1,pobl
    suma=suma+(partido(i)-r*total(i))**2
  end do
  var=(1.-n/pobl)*(1./n)*(pobl/tot)**2*suma/(pobl-1)
end subroutine

subroutine pptcr(partido,total,pi,n,r,var) !PPT con reemplazo
implicit none
integer:: n,i
integer, parameter::double=selected_real_kind(p=13)
real(kind=double)::r,var,votos,tot,suma,sumax,sumay,sumaxy
integer, parameter::pobl=63445
real(kind=double),dimension(pobl):: partido,total,pi
votos=0.
tot=0.
  do i=1,pobl
    !Calculando la razón muestral
    votos=votos+partido(i) !El valor poblacional de la razón
    tot=tot+total(i)
  end do
  r=votos/tot !Parámetro poblacional
  suma=0
  do i=1,pobl
    !Calculando la varianza del estimador
    suma=suma+(1./pi(i))*(partido(i)-r*total(i))**2
  end do
  var=(1./(n*tot**2))*suma
  sumax=0
  sumay=0
  sumaxy=0
  do i=1,pobl
    !Calculando la varianza del estimador
    sumax=sumax+pi(i)*(partido(i)/pi(i)-votos)**2
    sumay=sumay+pi(i)*(total(i)/pi(i)-tot)**2
    sumaxy=sumaxy+pi(i)*(partido(i)/pi(i)-votos)*
      (total(i)/pi(i)-tot)
  end do
  var=(1./(n*tot**2))*(sumax+r**2*sumay-2*r*sumaxy)
end subroutine

subroutine pptr(partido,total,pi,n,r,var) ! PPT sin reemplazo
implicit none
integer:: n,i
integer, parameter::double=selected_real_kind(p=13)
real(kind=double)::r,var,votos,tot,suma,sumax,sumay,sumaxy
integer, parameter::pobl=63445
real(kind=double),dimension(pobl):: partido,total,pi
votos=0.
tot=0.
  do i=1,pobl
    !Calculando la razón muestral
    votos=votos+partido(i) !El valor poblacional de la razón
    tot=tot+total(i)
  end do
  r=votos/tot !Parámetro poblacional
  suma=0

```



```

!      do i=1,pobl      !Calculando la varianza del estimador
!      suma=suma+(1./pi(i))*(partido(i)-r*total(i))**2
!      end do
!      var=(1./(n*tot**2))*suma
!      sumax=0
!      sumay=0
!      sumaxy=0
!      do i=1,pobl      !Calculando la varianza del estimador
!      sumax=sumax+pi(i)*(1-(n-1)*pi(i))*(partido(i)/pi(i)-votos)**2
!      sumay=sumay+pi(i)*(1-(n-1)*pi(i))*(total(i)/pi(i)-tot)**2
!      sumaxy=sumaxy+pi(i)*(1-(n-1)*pi(i))*(partido(i)/pi(i)-votos)*
+      (total(i)/pi(i)-tot)
!      end do
!      var=(sumax+r**2*sumay-2*r*sumaxy)/(n*tot**2)
!      end subroutine

```

A.2 Muestreo Estratificado, Estimador Separado y Combinado. Caso de Circunscripciones

```

program archivos4
implicit none
integer indica
integer, parameter::double=selected_real_kind(p=13)
integer i,n
integer, parameter::h=10
integer,dimension(h+1)::Nh,nhm,Nhacum
real(kind=double),dimension(63445)::di,se,li,pan,pri,prd,pcd,parm,
+ ds,nr,nulos,total,en,tipo,
+ circ,totalsvn,otros,dina,
+ entip,citip,tipo2,cirtip,etipo
real(kind=double)::pa_cam,ppri,pa_mex,va_cam,vpri,va_mex,
+ potros,votros
character(len=30)::nombre
character(len=35),dimension(63445)::mun
open(unit=5,file='resulta5',status='replace',action='write')
nombre='c:\ele2000\base5.dat'
open(unit=3,file=nombre,status='old',action='read',iostat=indica)
i=1
do ! Leyendo la base de datos
read(3,*,iostat=indica) di(i),se(i),li(i),pan(i),pri(i),prd(i),
+ pcd(i),parm(i),ds(i),nr(i),nulos(i),total(i),en(i),tipo(i),
+ circ(i),mun(i),totalsvn(i),otros(i),dina(i),citip(i),
+ entip(i),tipo2(i),cirtip(i),etipo(i)
if (indica /=0) exit
i=i+1
end do
close(unit=3)
! Muestreo estratificado el estrato es la circunscripcion y el
tipo de sección
do n=2000,2000
call estratos(cirtip,h,n,Nh,nhm,Nhacum)
call estrsepar(h,Nh,nhm,Nhacum,pan,total,pa_cam,va_cam)
write(5,703) n,pa_cam,va_cam
703 format(i4,' MESCT',' Alianza Cambio',5x,f7.4,5x,f14.10)
call estrsepar(h,Nh,nhm,Nhacum,pri,total,ppri,vpri)
write(5,803) n,ppri,vpri

```

```

803      format(i4, ' MESCT', 5x, ' PRI', 11x, f7.4, 5x, f14.10)
      call estrsepar(h, Nh, nhm, Nhacum, prd, total, pa_mex, va_mex)
      write(5, 903) n, pa_mex, va_mex
903      format(i4, ' MESCT', ' Alianza México', 5x, f7.4, 5x, f14.10)

      call estrsepar(h, Nh, nhm, Nhacum, otros, total, potros, votros)
      write(5, 953) n, potros, votros
953      format(i4, ' MESCT', 6x, ' Otros', 8x, f7.4, 5x, f14.10)
      call estrcomb(h, Nh, nhm, Nhacum, pan, total, pa_cam, va_cam)
      write(5, 1003) n, pa_cam, va_cam
1003     format(i4, ' MECCT', ' Alianza Cambio', 5x, f7.4, 5x, f14.10)
      call estrcomb(h, Nh, nhm, Nhacum, pri, total, ppri, vpri)
      write(5, 1103) n, ppri, vpri
1103     format(i4, ' MECCT', 5x, ' PRI', 11x, f7.4, 5x, f14.10)
      call estrcomb(h, Nh, nhm, Nhacum, prd, total, pa_mex, va_mex)
      write(5, 1203) n, pa_mex, va_mex
1203     format(i4, ' MECCT', ' Alianza México', 5x, f7.4, 5x, f14.10)

      call estrcomb(h, Nh, nhm, Nhacum, otros, total, potros, votros)
      write(5, 1253) n, potros, votros
1253     format(i4, ' MECCT', 4x, ' Otros', 8x, f7.4, 5x, f14.10)

      enddo
      close(5)
      end program

      subroutine estratos(estrat, h, n, Nh, nhm, Nhacum)
      implicit none
      !Subrutina que cuenta el número de secciones por estrato
      !y asigna proporcionalmente el tamaño de muestra por estrato
      integer, parameter::double=selected_real_kind(p=13)
      integer, parameter::pobl=63445
      real(kind=double), dimension(pobl)::estrat
      integer h, k, j, n
      integer, dimension(h+1)::Nh, nhm, Nhacum
      do k=1, h+1
         Nh(k)=0
         nhm(k)=0
         Nhacum(k)=0
      enddo
      do k=1, h      !Contando secciones por estrato
         do j=1, pobl
            if (estrat(j)==k) then
               Nh(k)=Nh(k)+1
            endif
         enddo
         nhm(k)=anint(1.*Nh(k)*n/pobl+1.) !tamaño de muestra por estrato,
         !asignación proporcional
      enddo
      do k=2, h+1
         Nhacum(k)=Nhacum(k-1)+Nh(k-1)      !secciones de estratos
      enddo
      enddo
      end subroutine

      subroutine estrsepar(h, Nh, nhm, Nhacum, partido, total, razon, var)
      implicit none

```



```

!Subrutina que calcula la varianza del estimador de razón
!utilizando el diseño estratificado separado
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::partido, total
integer i,h,k,j
integer,dimension(h)::Nh,nhm,Nhacum
real,dimension(h)::R,sumah
real(kind=double)::Xh,Yh,var,razon,Xt,Yt
Xt=0
Yt=0
do j=1, pobl
  Xt=Xt+partido(j)
  Yt=Yt+total(j)
enddo
Razon=Xt/Yt !Razón poblacional
do k=1,h
  Xh=0.
  Yh=0.
  do j=Nhacum(k)+1,Nhacum(k+1)
    Xh=Xh+partido(j)
    Yh=Yh+total(j)
  enddo
  R(k)=Xh/Yh !Razón por estrato
enddo
do k=1,h
enddo
!Se va a calcular la varianza del estimador usando la fórmula XX
var=0
do k=1,h
  sumah(k)=0.

  do i=1+Nhacum(k),Nhacum(k+1)
    sumah(k)=sumah(k)+(partido(i)-R(k)*total(i))**2
  enddo
  var=var+(1.*Nh(k)/pobl*1.)**2*(1-1.*nhm(k)/Nh(k))*1./nhm(k)*
+ 1./(Nh(k)-1)*sumah(k)
enddo
var=var/(1.*Yt/pobl)**2
end subroutine

```

```

subroutine estrcomb(h,Nh,nhm,Nhacum,partido,total,r,var)
implicit none
!Subrutina que calcula la varianza del estimador de razón
!utilizando el diseño estratificado combinado
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::partido, total
integer i,h,k,j
integer,dimension(h+1)::Nh,nhm,Nhacum
real,dimension(h+1)::sumah,Xhmedia,Yhmedia
real(kind=double)::Xt,Yt,var,R,Xh,Yh
Xt=0.
Yt=0.
do j=1, pobl
  Xt=Xt+partido(j)

```




```

    read(3,*,iostat=indica)di(i),se(i),li(i),pan(i),pri(i),prd(i),
+   pcd(i),parm(i),ds(i),nr(i),nullos(i),total(i),en(i),tipo(i),
+   circ(i),mun(i),totalsvn(i),otros(i),dina(i),citip(i),
+   entip(i),tipo2(i),cirtip(i),etipo(i)
    if (indica /=0) exit
    i=i+1
end do
close(unit=3)

do n=10,300,10
do muestra=2,5
do i=1, Np
    m(i)=muestra
enddo
    ! Muestreo estratificado el estrato es el distrito
    call conglo(dina,Np,Mp,Mpcum)
    call bietap(Np,n,Mp,m,Mpcum,pan,total,pa_cam,var1,var2,va_cam)
701 write(5,701) n,muestra,pa_cam,va_cam,var1,var2
+   format(i4,2x,i4,' Alianza Cambio',5x,f7.4,5x,
+   f14.10,5x,f14.10,5x,f14.10)
    call bietap(Np,n,Mp,m,Mpcum,pri,total,ppri,var1,var2,vpri)
801 write(5,801) n,muestra,ppri,vpri,var1,var2
+   format(i4,2x,i4,5x,' PRI',11x,f7.4,5x,f14.10,5x,f14.10,
+   5x,f14.10,5x,f14.10)
    call bietap(Np,n,Mp,m,Mpcum,prd,total,pa_mex,var1,var2,va_mex)
901 write(5,901) n,muestra,pa_mex,va_mex,var1,var2
+   format(i4,2x,i4,' Alianza Mexico',5x,f7.4,5x,f14.10,5x,f14.10
+   ,5x,f14.10)
    call bietap(Np,n,Mp,m,Mpcum,otros,total,potros,var1,var2,votros)
951 write(5,951) n,muestra,potros,votros,var1,var2
+   format(i4,2x,i4,6x,' Otros',8x,f7.4,5x,f14.10,5x,
+   f14.10,5x,f14.10)

enddo
enddo
close(5)
end program

```

```

subroutine conglo(cong,Np,Mp,Mpcum)
implicit none
!Subrutina que cuenta el # de elementos en cada conglomerado
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::cong
integer Np,i,j
integer,dimension(Np+1)::Mp,Mpcum

do i=1,Np+1
    Mp(i)=0
    Mpcum(i)=0
enddo
do i=2,Np+1
do j=1,pobl
    if (cong(j)==i-1) then
        Mp(i)=Mp(i)+1
    endif

```

```

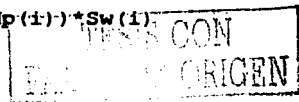
    enddo
  enddo
  mpcum(2)=mp(2)
  do i=2,Np+1
    mpcum(i)=mpcum(i-1)+mp(i)
  enddo
  do i=2,Np+1
    Mp(i-1)=Mp(i)
  enddo
end subroutine

subroutine bietap(Np,n,Mp,m,Mpcum,partido,total,R,var1,var2,var)
implicit none
!Subrutina que calcula varianza del estimador de
!razón en el muestreo bietápico. Ambas muestras se toman
!aleatorias simples
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::partido, total
integer i,Np,j,n
integer,dimension(Np+1)::Mp,Mpcum
integer,dimension(Np)::m
real,dimension(Np)::Sw,Xi,Yi,Ximedia,Yimedia
real(kind=double)::var,R,Xt,Yt,suma1,suma2,var1,var2
Xt=0
Yt=0
do j=1, pobl
  Xt=Xt+partido(j)
  Yt=Yt+total(j)
enddo
R=Xt/Yt
do i=1,Np
  Xi(i)=0.
  Yi(i)=0.
  do j=Mpcum(i)+1,Mpcum(i+1)
    Xi(i)=Xi(i)+partido(j)
    Yi(i)=Yi(i)+total(j)
  enddo
  Ximedia(i)=Xi(i)/Mp(i)      !medias en cada conglomerado
  Yimedia(i)=Yi(i)/Mp(i)
enddo
!Se va a calcular la varianza del estimador

suma1=0
suma2=0
do i=1,Np
  Sw(i)=0.

  do j=1+Mpcum(i),Mpcum(i+1)
    Sw(i)=Sw(i)+(partido(j)-Ximedia(i)-R*(total(j)-Yimedia(i)))**2
  enddo
enddo
do i=1,Np
  Sw(i)=Sw(i)/(Mp(i)-1)
  suma1=suma1+(Xi(i)-R*Yi(i))**2
  suma2=suma2+(1.*Mp(i)**2/m(i))*(1-1.*m(i)/Mp(i))*Sw(i)

```



```

enddo
var1=(1-1.*n/Np)*Np**2*(1./(n*Yt**2))*(1./(Np-1))*suma1
var2=Np*suma2/(n*Yt**2)
var=var1+var2
end subroutine

```

A.4 Muestreo Bietápico, Distritos Electorales como UPM y Secciones como USM. Muestreo PPT con reemplazo en la primera etapa y MAS en la segunda.

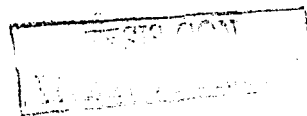
```

program bietpptmas
! Muestreo Bietápico, PPT sin reemplazo en UPM MAS en USM
! Ecuación 1.9, capítulo 1
implicit none
integer indica
integer, parameter::double=selected_real_kind(p=13)
integer i,muestra,n
integer, parameter::Np=300 !número de unidades 1a. etapa (Np=N)
! integer, parameter::n=300 !n=#conglomerados muestreados de los Np
integer, dimension(Np+1)::Mp,Mpcum
real, dimension(Np)::pi
integer, dimension(Np)::m !Mp(i)=#elementos en el congl.i
!m(i)= #elementos muestreados
en el congl.i

real(kind=double), dimension(63445)::di, se, li, pan, pri, prd, pcd, parm,
+ ds, nr, nulos, total, en, tipo,
+ circ, totalsvn, otros, dina,
+ entip, citip, tipo2, cirtip, etipo
real(kind=double)::pa_cam, ppri, pa_mex, va_cam, vpri, va_mex,
+ potros, votros, var1, var2
character(len=30)::nombre
character(len=35), dimension(63445)::mun
open(unit=5, file='resulta8', status='replace', action='write')
nombre='c:\ele2000\base3.dat'
open(unit=3, file=nombre, status='old', action='read', iostat=indica)
i=1
do ! Leyendo la base de datos
read(3, *, iostat=indica) di(i), se(i), li(i), pan(i), pri(i), prd(i),
+ pcd(i), parm(i), ds(i), nr(i), nulos(i), total(i), en(i), tipo(i),
+ circ(i), mun(i), totalsvn(i), otros(i), dina(i), citip(i),
+ entip(i), tipo2(i), cirtip(i), etipo(i)
if (indica /=0) exit
i=i+1
end do
close(unit=3)

do n=30, 30
do muestra=5, 5
do i=1, Np
m(i)=muestra
enddo
! Las UPM son los distritos
call conglo(dina, Np, Mp, Li, pi, Mpcum)

```




```

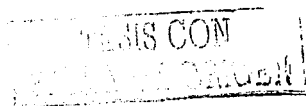
end subroutine

subroutine bietap(Np,n,Mp,pi,m,Mpcum,partido,total,R,var1,var2,
+
var)
implicit none
!Subrutina que calcula varianza del estimador de
!razón en el muestreo bietápico. PPT con reemplazo en UPM y
!MAS sin reemplazo en USM
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::partido, total
integer i,Np,j,n
integer,dimension(Np+1)::Mp,Mpcum
integer,dimension(Np)::m
real,dimension(Np)::Sw,Xi,Yi,Ximedia,Yimedia,pi
real(kind=double)::var,R,Xt,Yt,suma1,suma2,var1,var2
Xt=0
Yt=0
do j=1, pobl
Xt=Xt+partido(j)
Yt=Yt+total(j)
enddo
R=Xt/Yt

do i=1,Np
Xi(i)=0.
Yi(i)=0.
do j=Mpcum(i)+1,Mpcum(i+1)
Xi(i)=Xi(i)+partido(j)
Yi(i)=Yi(i)+total(j)
enddo
Ximedia(i)=Xi(i)/Mp(i) !medias en cada conglomerado
Yimedia(i)=Yi(i)/Mp(i)
enddo
!Se va a calcular la varianza del estimador
suma1=0
suma2=0
do i=1,Np
Sw(i)=0.

do j=1+Mpcum(i),Mpcum(i+1)
Sw(i)=Sw(i)+(partido(j)-Ximedia(i)-R*(total(j)-Yimedia(i)))**2
enddo
enddo
do i=1,Np
Sw(i)=Sw(i)/(Mp(i)-1)
suma1=suma1+1/pi(i)*(Xi(i)-R*Yi(i))**2
suma2=suma2+1/pi(i)*(1.*Mp(i)**2/m(i))*(1-1.*m(i)/Mp(i))*Sw(i)
enddo
var1=(1./(n*Yt**2))*suma1
var2=(1./(n*Yt**2))*suma2
var=var1+var2
end subroutine

```




```

801      write(5,801) n,muestra,ppri,vpri,var1,var2
      format(i4,2x,i4,5x,' PRI',11x,f7.4,5x,f14.10,5x,f14.10,
+         5x,f14.10,5x,f14.10)
      call bietap(Np,n,Mp,pi,m,Mpcum,prd,total,pa_mex,var1,var2,
+         va_mex)
901      write(5,901) n,muestra,pa_mex,va_mex,var1,var2
      format(i4,2x,i4,' Alianza Mexico',5x,f7.4,5x,f14.10,5x,f14.10
+         ,5x,f14.10)
      call bietap(Np,n,Mp,pi,m,Mpcum,otros,total,potros,var1,var2,
+         votros)
951      write(5,951) n,muestra,potros,votros,var1,var2
      format(i4,2x,i4,6x,' Otros',8x,f7.4,5x,f14.10,5x,
+         f14.10,5x,f14.10)

```

```

enddo
enddo
close(5)
end program

```

```

subroutine conglo(cong,Np,Mp,Li,pi,Mpcum)
implicit none
!Subrutina que cuenta el # de secciones y
!la lista nominal en cada conglomerado (distrito)
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::cong,li
integer Np,i,j
integer,dimension(Np+1)::Mp,Mpcum,Lp
real,dimension(Np)::pi
do i=1,Np+1
  Mp(i)=0
  Lp(i)=0
  Mpcum(i)=0
enddo
do i=2,Np+1
  do j=1,pobl
    if (cong(j)==i-1) then
      Mp(i)=Mp(i)+1
      Lp(i-1)=Lp(i-1)+li(j)
    endif
  enddo
enddo
mpcum(2)=mp(2)
do i=2,Np+1
  mpcum(i)=mpcum(i-1)+mp(i)
enddo
do i=2,Np+1
  Mp(i-1)=Mp(i)
  pi(i-1)=1.0*Mp(i-1)/pobl
enddo
end subroutine

```

```

subroutine bietap(Np,n,Mp,pi,m,Mpcum,partido,total,R,var1,var2,
+         var)
implicit none
!Subrutina que calcula varianza del estimador de

```

```

!razón en el muestreo bietápico. PPT con reemplazo en UPM y
!MAS sin reemplazo en USM
integer, parameter::double=selected_real_kind(p=13)
integer, parameter::pobl=63445
real(kind=double), dimension(pobl)::partido, total
integer i, Np, j, n
integer, dimension(Np+1)::Mp, Mpcum
integer, dimension(Np)::m
real, dimension(Np)::Sw, Xi, Yi, Ximedia, Yimedia, pi
real(kind=double)::var, R, Xt, Yt, suma1, suma2, var1, var2
Xt=0
Yt=0
do j=1, pobl
  Xt=Xt+partido(j)
  Yt=Yt+total(j)
enddo
R=Xt/Yt

do i=1, Np
  Xi(i)=0.
  Yi(i)=0.
  do j=Mpcum(i)+1, Mpcum(i+1)
    Xi(i)=Xi(i)+partido(j)
    Yi(i)=Yi(i)+total(j)
  enddo
  Ximedia(i)=Xi(i)/Mp(i)      !medias en cada conglomerado
  Yimedia(i)=Yi(i)/Mp(i)
enddo
!Se va a calcular la varianza del estimador
suma1=0
suma2=0
do i=1, Np
  Sw(i)=0.

  do j=1+Mpcum(i), Mpcum(i+1)
    Sw(i)=Sw(i)+(partido(j)-Ximedia(i)-R*(total(j)-Yimedia(i)))**2
  enddo
enddo
do i=1, Np
  Sw(i)=Sw(i)/(Mp(i)-1)
  suma1=suma1+1/pi(i)*(Xi(i)-R*Yi(i))**2
  suma2=suma2+1/pi(i)*(1.*Mp(i)**2/m(i))*(1-1.*m(i)/Mp(i))*Sw(i)
enddo
var1=(1./(n*Yt**2))*suma1
var2=(1./(n*Yt**2))*suma2
var=var1+var2
end subroutine

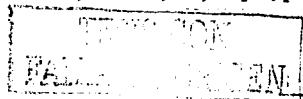
```

A.5 Sesgo y Coeficiente de variación en el Muestreo Aleatorio Simple

```

program massesgo
implicit none
integer indica
integer, parameter::double=selected_real_kind(p=13)
integer i, n
real(kind=double), dimension(63445)::di, se, li, pan, pri, prd, pcd, parm,
+
ds, nr, nulos, total, en, tipo, pi,

```



```

+
circ,totalsvn,otros,dina,
+
entip,ctip,tipo2,cirtip,etipo
real(kind=double)::span,spri,sprd,total,sotros,sesgo,cvy
real(kind=double)::pa_cam,ppri,pa_mex,va_cam,vpri,va_mex,
+
potros,votros
character(len=30)::nombre1,nombre2
character(len=35),dimension(63445)::mun
nombre1='c:\ele2000\resulta9.dat'
open(unit=7,file=nombre1,status='replace',action='write')
nombre2='c:\ele2000\base1.dat'
open(unit=3,file=nombre2,status='old',action='read',iostat=indica)
span=0.
spri=0.
sprd=0.
stotal=0.
sotros=0.
i=1
do ! Leyendo la base de datos
read(3,*,iostat=indica)di(i),se(i),li(i),pan(i),pri(i),prd(i),
+
pcd(i),parm(i),ds(i),nr(i),nulos(i),total(i),en(i),tipo(i),
+
circ(i),mun(i),totalsvn(i),otros(i),dina(i),citip(i),
+
entip(i),tipo2(i),cirtip(i),etipo(i)
if(indica/=0)exit
i=i+1
enddo
close(unit=3)
do i=1,63445 !Calculando tamaños relativos por sección
pi(i)=li(i)/58760114.
enddo

do n=300,1500,300
! Muestreo Aleatoiro Simple

call mas(pan,total,n,pa_cam,va_cam,sesgo,cvy)
write(7,700) n,pa_cam,va_cam,sesgo,cvy
700 format(i4,'MAS ',7x,'Alianza Cambio',1x,f7.4,5x,f14.10,
+
5x,f14.10,5x,f14.10)
call mas(pri,total,n,ppri,vpri,sesgo,cvy)
write(7,800) n,ppri,vpri,sesgo,cvy
800 format(i4,'MAS ',7x,'PRI',12x,f7.4,5x,f14.10,
+
5x,f14.10,5x,f14.10)
call mas(prd,total,n,pa_mex,va_mex,sesgo,cvy)
write(7,900) n,pa_mex,va_mex,sesgo,cvy
900 format(i4,'MAS ',7x,'Alianza México',1x,f7.4,5x,f14.10,
+
5x,f14.10,5x,f14.10)
call mas(otros,total,n,potros,votros,sesgo,cvy)
write(7,950) n,potros,votros,sesgo,cvy
950 format(i4,'MAS ',7x,'Otros',10x,f7.4,5x,f14.10,
+
5x,f14.10,5x,f14.10)
enddo
close(7)
end

subroutine mas(partido,total,n,r,var,sesgo,cvy)
implicit none

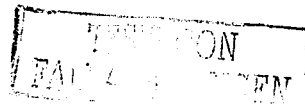
```



```

integer::i,n
integer, parameter::double=selected_real_kind(p=13)
real(kind=double)::r,var,votos,tot,sesgo,sy,sxy,mediay,var2,cvy
+ ,muestra
real ::suma
integer, parameter::pobl=63445
real(kind=double),dimension(pobl)::partido,total
votos=0.
tot=0.
do i=1,pobl
    votos=votos+partido(i)
    tot=tot+total(i)
end do
r=votos/tot
mediay=1.*tot/pobl
suma=0
sy=0
sxy=0
do i=1,pobl
    suma=suma+(partido(i)-r*total(i))**2
    sy=sy+(total(i)-tot/pobl)**2
    sxy=sxy+(partido(i)-votos/pobl)*(total(i)-tot/pobl)
end do
sesgo=(1./(pobl-1))*(1.-n/pobl)*(1./n)*1./(tot/pobl)**2*(r*sy-sxy)
var=(1.-n/pobl)*(1./n)*(pobl/tot)**2*suma/(pobl-1)
var2=(1.-n/pobl)*(1./n)*(1./(pobl-1))*sy
cvy=sqrt(var2)/mediay
sy=sy/(pobl-1)
muestra=(1.*100*pobl*sy/mediay**2)/(pobl+100*sy/mediay**2)
write(*,*) muestra
end subroutine

```



Referencias

- Cochran, W. G.
1993 *Técnicas de Muestreo*. CECSA, México, Traducción de Andrés Sestier B.
- Eslava G., I. Méndez and P. Romero
Sample designs for electoral surveys in Mexico. Ponencia presentada en la Conferencia de la Asociación Internacional de Estadísticas Oficiales (IAOS), Montreux, Suiza 4-8 septiembre 2000
- Hartley, H. O. y Rao J. N. K.
1962 Sampling With Unequal Probabilities and Without Replacement. *Annals of Mathematical Statistics*, 33, 350-374
- Instituto Federal Electoral
Página electrónica: www.ife.org.mx
- Instituto Federal Electoral
Atlas Electoral Federal de México 1991-2000, Dirección Ejecutiva de Organización Electoral, version cd-room
- Kish, Leslie
1975 *Muestreo de Encuestas*. Trillas, México, Traducción de Ricardo Vinós Cruz López
-
- 1987 *Statistical Design For Research*. John Wiley and Sons, New York
- Raj Des
1980 Teoría del Muestreo. Fondo de Cultura Económica, México. Traducción de Roberto R. Reyes-Mazzoni

TESIS CON
FALLA DE ORIGEN