



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MEXICO

RECIBIDA
ACATLAN
2023
14

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
ACATLAN

EL MUESTREO PARA PROPORCIONES. UN CASO
PRACTICO EN ENCUESTAS PARA MEDICION DE
PREFERENCIAS ELECTORALES

T E S I S
QUE PARA OBTENER EL TITULO DE
ACTUARIA
PRESENTA
NORMA LEYVA ESTRADA



ASESOR: Mtro. RICARDO CESAR APARICIO JIMENEZ



ACATLAN, EDO. DE MEX.

1995

FALLA DE ORIGEN



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 está 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.



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLAN"

DIVISION DE MATEMATICAS E INGENIERIA
PROGRAMA DE ACTUARIA Y M.A.C.

SRITA. NORMA LEYVA ESTRADA
Alumna de la carrera de Actuaría
P r e s e n t e .

Por acuerdo a su solicitud presentada con fecha 28 de septiembre de 1994, me complace notificarle que esta Jefatura tuvo a bien asignarle el siguiente tema de tesis: "EL MUESTREO PARA PROPORCIONES, UN CASO PRACTICO EN ENCUESTAS PARA MEDICION DE PREFERENCIAS ELECTORALES", el cual se desarrollará como sigue:

INTRODUCCION

- CAP. I Aspectos teóricos del muestreo para proporciones.
 - CAP. II Revisión empírica de factores que influyen en la precisión de estimaciones de proporciones.
 - CAP. III Aplicación práctica del muestreo para proporciones en la medición de preferencias electorales.
 - CAP. IV Análisis y presentación de resultados.
- CONCLUSIONES

Asimismo, fue designado como Asesor de Tesis el MTR. RICARDO CESAR APARICIO JIMENEZ, Profesor de esta Escuela.

Ruego a usted tomar nota que en cumplimiento de lo especificado en la Ley de Profesiones, deberá presentar servicio social durante un tiempo mínimo de seis meses - como requisito básico para sustentar examen profesional así como de la disposición de la Coordinación de la Administración Escolar en el sentido de que se imprima en lugar visible de los ejemplares de la Tesis el título del trabajo realizado. Esta comunicación deberá imprimirse en el interior de la misma.

A T E N T A M E N T E
"POR MI RAZA HABLARA EL ESPIRITU"
Acatlán, D.F. Méx. enero 20 de 1995.

ACT. LAURA ANA NAVERA BECERRA
Jefe del Programa de Actuaría
y M.A.C.

cg'



*A mis papás,
a Yola, Lulú y Efrén
con todo mi cariño*

AGRADECIMIENTOS

En la elaboración de este trabajo de tesis tuve la oportunidad de contar con la colaboración de todo un grupo de amigos. En este espacio quiero brindarles mi más sincero agradecimiento por su tiempo e interés.

Agradezco a...

Mis papás... su amor, guía, apoyo y comprensión

Yola, Lulú y Frenchas... ser mis amigos, compañeros y hermanos

Helen... el haber sido la chispa de impulso y asesora de este trabajo

Ricardo... su tiempo para asesorarme y paciencia y comprensión ante mis titubeos

Moni... su ayuda incondicional

Lauris, Jesús y José Manuel... su ayuda para no perder el objetivo

Compañeros de la escuela... 4 años de convivencia, de compartir experiencias y sueños

Profesores... su "amor a la camiseta"

Reconozco y hago explícito que en gran parte este escrito pudo llegar a feliz término gracias a ustedes.

De manera muy especial quiero agradecer al Dr. Ricardo Vernon Carter, Director General de Opinión Profesional S. A. de C. V., por el apoyo y las facilidades que me brindó para poder realizar este estudio.

CONTENIDO

INTRODUCCION

1.	ASPECTOS TEORICOS DEL MUESTREO PARA PROPORCIONES	1
1.1.	¿Qué son las proporciones?	1
1.2.	Estimación de proporciones a partir de una muestra	2
1.3.	Cálculo del tamaño de muestra para una proporción	9
1.4.	Estimación de proporciones a partir de diferentes diseños de muestreo	14
1.5.	Efecto de diseño	19
1.6.	Precisión en la estimación de proporciones	21
2.	REVISION EMPIRICA DE FACTORES QUE INFLUYEN EN LA PRECISION DE ESTIMACIONES DE PROPORCIONES	25
2.1.	Tamaño de muestra	25
2.2.	Tamaño de muestra para subpoblaciones	31
2.3.	Varianza y presencia de la característica de interés en la población	33
3.	APLICACION PRACTICA DEL MUESTREO PARA PROPORCIONES EN LA MEDICION DE PREFERENCIAS ELECTORALES	39
3.1.	Mecanismos de obtención de información electoral	39
3.1.1.	Acopio inmediato de resultados electorales (AIRE)	41
3.1.2.	Encuesta Posvoto (EP)	42
3.2.	Metodología del Acopio Inmediato de Resultados Electorales	44
3.2.1.	Diseño de muestreo	45
3.2.2.	Operativo de campo	49
3.2.3.	Procesamiento de la información	50
3.3.	Metodología de la Encuesta Posvoto	54
3.3.1.	Diseño de muestreo	54
3.3.2.	Operativo de campo	56
3.3.3.	Procesamiento de la información	58
4.	ANALISIS Y PRESENTACION DE RESULTADOS	62
4.1.	Análisis estadístico de algunos operativos de AIRE	62
4.2.	El alcance de una Encuesta Posvoto	76
	CONCLUSIONES	88
	BIBLIOGRAFIA	93

ANEXO A. Fórmulas para la estimación de proporciones y sus varianzas a partir de diferentes diseños de muestreo	95
ANEXO B. Tablas de valores de los intervalos de confianza para las estimaciones de los operativos de AIRE	99
ANEXO C. Diferencia de proporciones. Prueba ji-cuadrada. Fórmulas y cálculos	104
ANEXO D. Cuestionario de la Encuesta Posvoto en Yucatán	109

INTRODUCCION

El muestreo es una técnica especializada dentro de la estadística y se aplica en áreas de investigación científica, social y comercial. Entre las medidas cuantitativas que se utilizan para presentar los resultados generados de una encuesta por muestreo, la media aritmética es muy común. Un caso particular de ésta, es la proporción, ya que equivale a calcular la media de una variable que sólo puede tomar dos valores: uno cuando posee y cero cuando no posee una característica de interés.

El propósito de este trabajo es fundamentar de manera empírica algunos de los factores que influyen en la precisión de las estimaciones de proporciones generadas a través de encuestas por muestreo y ver, en la aplicación práctica, el cumplimiento de conceptos teóricos del muestreo para realizar este tipo de estimaciones. Esto debido a que las proporciones y los porcentajes son conceptos de fácil comprensión, ya que dan una idea clara de la presencia de cierta característica de interés en la población de estudio, y porque se usan de manera frecuente en la presentación de resultados. En este estudio se piensa al muestreo para proporciones como una herramienta útil y necesaria para que se pueda llevar a cabo la planeación, el procesamiento y el análisis de encuestas para medición de preferencias electorales.

Para desarrollar el tema se decidió escoger una encuesta real aplicada en México, en lugar de un ejemplo hipotético. Dado el momento político que vivía el país al momento de hacer el trabajo, se seleccionaron encuestas que buscaban medir las preferencias electorales la noche misma del día de las elecciones.

Respecto a las encuestas para la medición de preferencias electorales que se utilizan en este estudio, no se pretende agotar el análisis estadístico que se puede realizar siendo encuestas

de las que puede surgir mucha información valiosa, sino que como se enuncia anteriormente, ver la aplicación práctica del muestreo para proporciones. Además, la estructura del trabajo permite que, tanto personas interesadas en el muestreo como personas con conocimientos básicos en estadística que busquen conocer más sobre este tipo de encuestas, puedan comprender la importancia de la teoría del muestreo.

Este trabajo consta de cuatro capítulos. En el primer capítulo se presentan los conceptos teóricos básicos acerca del muestreo para proporciones, la estimación de una proporción a partir de una muestra, el cálculo del tamaño de una muestra cuando se busca estimar proporciones, los diferentes tipos de diseños de muestreo básicos, el efecto del diseño de muestreo en las estimaciones y la precisión de las mismas. Para tratar los conceptos que se acaban de mencionar, se toma como base la aplicación del muestreo para proporciones bajo el supuesto de una muestra seleccionada con muestreo aleatorio simple, diseño de muestreo considerado como el camino más elemental para calcular estadísticas a partir de muestras. Existe bibliografía amplia destinada específicamente al tratamiento de la teoría de muestreo, por lo que este capítulo no pretende agotar el tema, sino exponer los conceptos básicos para la comprensión general del trabajo.

El segundo capítulo busca presentar, a partir de un estudio empírico, la importancia del tamaño de la muestra, la varianza de los estimadores y el nivel de presencia del fenómeno bajo estudio en la precisión de estimaciones de proporciones. Como en muchas áreas del conocimiento humano, en el muestreo los conceptos no son independientes sino que están interrelacionados, por lo que también se trata en este capítulo el uso del factor de corrección para poblaciones finitas, la distribución normal y el efecto de los errores de muestreo.

La aplicación del muestreo para proporciones es el objetivo del tercer capítulo, enfocado en particular a la medición de preferencias electorales. Los dos tipos de encuestas que se utilizan se aplican en México y en diversas partes del mundo para obtener estimaciones de

resultados electorales: el Acopio Inmediato de Resultados Electorales (AIRE) y la Encuesta Posvoto (EP). En términos generales, la primera consta de resultados de la votación real de las casillas en muestra y la segunda, de entrevistas realizadas a votantes a la salida de las casillas, en una muestra de secciones previamente seleccionadas. En este capítulo se explican más a detalle los objetivos y el operativo de campo de cada una de estas encuestas, así como la metodología y el procesamiento de la información aplicados en México por una empresa mexicana, Opinión Profesional S. A. de C. V., que a la fecha ha realizado operativos de AIRE y EP para 63 y 34 dominios de estudio, respectivamente.

El cuarto capítulo abarca un análisis de la información que se genera a partir de cada uno de los dos tipos de encuestas mencionados, y se observa el cumplimiento de algunos conceptos teóricos del muestreo para proporciones. Primero se analizan los diferentes niveles de presencia de la característica de interés, en este caso preferencia electoral, a partir de levantamientos de AIRE y se observa el comportamiento de las estimaciones en cada uno de los niveles: predominio de un partido, competencia entre dos partidos o más y presencia rara de un partido. La segunda sección de este capítulo se ocupa de la información que se puede generar a partir de una EP adicional a la posibilidad de estimar la distribución de los votos.

1. ASPECTOS TEORICOS DEL MUESTREO PARA PROPORCIONES

Este capítulo tiene como objetivo presentar el manejo de las proporciones en la teoría del muestreo, por lo que el contenido se enfoca a explicar los conceptos básicos al respecto. Primero, se da una breve explicación del uso de una proporción y su cálculo en un muestreo aleatorio simple. Además, se especifica la importancia de decidir el tamaño de muestra necesario para la estimación de una proporción. En seguida, se plantea la forma de estimar proporciones a partir de diferentes diseños de muestreo. Así, una vez especificados algunos tipos de diseños de muestreo, se trata el efecto de diseño en las estimaciones. Finalmente, siendo la precisión de las estimaciones obtenidas con una muestra un aspecto fundamental, se dedica una sección para su tratamiento.

1.1. ¿QUE SON LAS PROPORCIONES?

La tendencia en el incremento en el uso de información y a su vez en el análisis e interpretación de resultados, señala la necesidad de disponer de medidas cuantitativas que representen el fenómeno de estudio. Una de ellas es la media, una medida de localización que se conoce comúnmente como "promedio". La media de un conjunto de n números es simplemente la división de la suma de la característica de interés por n .

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

Un caso particular de esta media aritmética es la proporción. Una proporción p es la media de una variable dicotómica Y_i (variable que sólo puede tomar dos valores diferentes), en

donde los miembros que poseen un atributo reciben el valor de $y_i = 1$, mientras que los que no poseen dicho atributo tienen el valor de $y_i = 0$.

Es fácil demostrar que una proporción sólo puede tomar valores entre 0 y 1, por lo que se acostumbra más el uso de porcentajes, los cuales se obtienen multiplicando la proporción respectiva por 100.

En particular, muchos de los resultados de encuestas se presentan en porcentajes; por ejemplo, el $x\%$ de la población mexicana que es mayor de cierta edad, $x\%$ de las empresas mexicanas pertenecen al sector manufacturero, $x\%$ de los países en el mundo son desarrollados, etc. Su uso constante y fácil comprensión han hecho de esta presentación de resultados una forma común de reporte.

Las proporciones y los porcentajes dan una idea clara de la presencia de cierta característica en la población de estudio; por tal motivo el presente trabajo se centra en su manejo.

1.2. ESTIMACION DE PROPORCIONES A PARTIR DE UNA MUESTRA

Una forma muy difundida de observación directa y obtención de información actual es la realización de encuestas. Tal vez, la necesidad de información acerca de aspectos de la vida cotidiana, referentes a una gran variedad de disciplinas del conocimiento humano, ha provocado el incremento en número y tipos de encuestas que se realizan.

Una gran parte de las encuestas que se levantan se basan en muestras, es decir, la información se recolecta alrededor de sólo una fracción de la población acerca de la cual se quiere hacer la inferencia, en lugar de obtenerla para cada miembro de la población. Esto permite tener costos menores, la utilización de los resultados puede realizarse en menos

tiempo al reducir las cargas de trabajo, la calidad de los resultados puede mejorarse y supervisarse, etc.

Precisamente, el objetivo fundamental de una de las dos ramas de la estadística, la estadística inferencial, es hacer inferencias acerca de una población con base en la información contenida en una muestra. Cabe señalar que las encuestas permiten hacer afirmaciones fidedignas sobre todos, sobre el grupo, sobre el agregado; pero no sobre cada cual, no sobre los individuos.

Las muestras se pueden seleccionar, ya sea de manera intencional, es decir, la persona que selecciona la muestra escoge, según su juicio, qué elementos entran en la muestra, o bien de manera probabilística, en donde la selección de la muestra puede considerarse un experimento aleatorio. Cabe mencionar que el hecho de que todas las unidades en la población tengan una probabilidad conocida y no nula de ser seleccionadas es lo que científicamente permite inferir los resultados de la muestra a toda la población.

Una parte importante del desarrollo de una encuesta es el marco o lista de muestreo. Consiste en el grupo de elementos que tienen la oportunidad de ser seleccionados, dada la aproximación de muestreo que se escoja. Por lo tanto, debe cuidarse que la cobertura de la población estudiada sea adecuada, ya que una muestra sólo puede ser representativa de la población incluida en el marco de muestreo.

Para el diseño de muestras existe una técnica especializada dentro de la estadística: el muestreo. Actualmente se aplica a muchas áreas de investigación científica, social y comercial.

El diseño de muestras tiene dos aspectos:

- "un proceso de selección, que consiste en las reglas y operaciones mediante las cuales se incluyen en la muestra algunos miembros de la población, y
- un proceso de estimación para calcular las estadísticas de la muestra, que son estimaciones muestrales de valores de la población"¹.

Referente al primer aspecto del diseño de muestras, existen diversos métodos de selección. Sin embargo, por el momento se va a trabajar con lo que se denomina muestreo aleatorio simple, el camino más elemental para calcular estadísticas a partir de muestras. Este puede verse como el proceso de selección base, siendo los demás procedimientos modificaciones de él. Por esta razón el muestreo aleatorio simple es el más adecuado para una introducción a la teoría del muestreo.

El muestreo aleatorio simple asume un tamaño de muestra n , seleccionada aleatoriamente de una población de tamaño N , en donde cada muestra posible de n elementos tiene la misma probabilidad de ser seleccionada. Para seleccionar la muestra se enumeran los elementos y se escogen generalmente con la ayuda de una tabla de números aleatorios o de algún programa de computación que genere números aleatorios.

En lo que se refiere a la notación que se va a utilizar, se debe aclarar que las letras mayúsculas tienen que ver con los valores de la población, y las letras minúsculas denotan los valores de la muestra. Así:

N	número de elementos en la población
Y_i	valor de la característica y para el elemento i -ésimo de la población
n	número de elementos en la muestra
y_i	valor de la característica y para el elemento i -ésimo de la muestra

¹ Kish, Leslie. *Muestreo de Encuestas*. Trillas. México, 1975. p. 24

Para estimar la media de una muestra seleccionada mediante muestreo aleatorio simple se ocupa el estimador

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

es decir, se suman los valores de la variable en estudio de cada uno de los elementos de la muestra y se divide por el número de elementos en la muestra.

Ahora, si esto se concibe para una proporción, el proceso de estimación consiste en sumar 1's ó 0's según el elemento presente o no la característica de interés, y dividir el resultado por el número de elementos en la muestra. Se observa entonces que el estimador de una proporción, está basado en el mismo principio que el de la media sólo que:

$y_i = 1$ si el i -ésimo elemento presenta la característica de interés, y
 $y_i = 0$ en otro caso

Si se denota por a a la suma de 1's, es decir, a la suma de valores de y_i de los elementos que tienen determinado atributo, entonces

$$p = \frac{a}{n}.$$

Además de la media, las mediciones en las encuestas por muestreo usualmente se enfocan a la estimación de ciertas características numéricas de la población, tales como el total poblacional Y .

Así, el estimador del agregado o total poblacional está dado por

$$\hat{Y} = y = N\bar{y}$$

y si la media es una proporción

$$\hat{Y} = Np.$$

Este último se refiere, de manera desagregada, a los elementos en la población total que corresponden a determinada proporción, en otras palabras, cuántas personas en una población presentan cierta característica.

Dados un proceso de selección de los elementos en la muestra y el tamaño de la misma, "al conjunto de los posibles valores de un estimador cada uno con su probabilidad de materializarse se le llama distribución de muestreo"², donde una muestra determinada representa un punto de la distribución de muestreo.

Para conocer su variabilidad se calcula su varianza. La varianza refleja qué tan dispersas están las mediciones del estimador, por eso se considera una medida de dispersión, en este caso, ya sea de la media o bien del total poblacional.

Cabe mencionar que "la medibilidad es una característica de los diseños que permite calcular, a partir de la muestra, estimaciones válidas o aproximaciones de su variabilidad de muestreo"³. Luego, es necesario obtener estimadores de varianza de los estimadores para poder hacer inferencias válidas.

La varianza de una población está dada por

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

² *Ibid.*, p. 31

³ *Ibidem.*, p. 45

Sin embargo, esta ecuación carece de uso práctico, ya que su valor depende de los valores de toda la población y, por tanto, no se puede calcular a partir de una muestra.

Otra forma de denotar la varianza de la población es

$$S^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}.$$

"Su ventaja es que la mayoría de los resultados toman una forma ligeramente más simple"⁴ pero, como la anterior, se requieren para su cálculo todos y cada uno de los elementos de la población.

Un estimador de S^2 basado en un muestreo aleatorio simple⁵ se puede obtener con la siguiente ecuación

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

Finalmente, lo que se calcula es la varianza de un estimador, digamos la media. Entonces, su varianza está dada por la siguiente ecuación:

$$\sigma_{\bar{y}}^2 = S_{\bar{y}}^2 = V(\bar{y}) = \frac{S^2}{n}(1-f),$$

donde $f = n/N$ es la fracción de muestreo y al factor $(1-f)$ se le llama factor de corrección para poblaciones finitas (*cpf*), y es simplemente la proporción de la población no incluida en la muestra.

⁴ Cochran, William G. *Técnicas de Muestreo*, C.E.C.S.A. México, 1980, p. 47

⁵ Dado que es una estimación realizada a partir de una muestra, debe considerarse el proceso de selección de la muestra.

Luego, para que se pueda obtener esta varianza a partir de una muestra, en lugar de S^2 se utiliza s^2 . Así, la varianza estimada de la media \bar{y} de una muestra aleatoria simple se calcula mediante⁶:

$$s_y^2 = v(\bar{y}) = \frac{s^2}{n}(1-f),$$

donde

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

y $f = n/N$.

Como puede observarse, a medida que N crece, la fracción de muestreo tiende a cero y el factor $(1-f)$ se aproxima a la unidad. Entonces la *cpf* puede omitirse cuando la población es mucho mayor que la muestra, digamos que cuando f no excede un 5%. En una población infinita la *cpf* desaparece de la fórmula de la varianza, de ahí su nombre.

El efecto de ignorar la corrección para poblaciones finitas es sobrestimar la varianza de la estimación \bar{y} .

La estimación de la varianza de una proporción está dada por:

$$v(p) = \frac{pq}{n-1}(1-f),$$

donde

$q = 1 - p$ denota la proporción que carece de la característica de interés.

Ya se ha visto que una proporción es simplemente un caso especial de una media. De esto se sigue que la ecuación de $v(p)$ se puede obtener a partir de la ecuación $v(\bar{y})$. Como se puede

⁶ Las letras minúsculas se ocupan para denotar estimadores.

observar, el cálculo de la varianza para proporciones es más simple. Dada la frecuencia de uso de este caso particular de la media, generalmente se trata en forma separada.

La estimación de la varianza del total será

$$v(\hat{Y}) = v(y) = N^2 v(\bar{y}).$$

La variabilidad en la estimación de totales en el muestreo aleatorio simple es por tanto, N^2 veces mayor que la de la media. Sin embargo, ya que las unidades de medición son diferentes en ambos casos, esta observación no tiene un efecto directo en qué tan "buena" sea una estimación.

Hasta este momento se tienen algunos conceptos básicos de los estimadores que suelen utilizarse para el manejo de la información recabada en encuestas por muestreo. En dichos estimadores se tiene una n dada. Sin embargo, no se ha mencionado nada acerca de cómo se designa el tamaño de esa muestra. En la siguiente sección se considerará este problema.

1.3. CALCULO DEL TAMAÑO DE MUESTRA PARA UNA PROPORCION

Existen dos formas comunes pero inapropiadas de calcular el tamaño de una muestra.

La primera idea errónea es pensar que el tamaño de una muestra depende fuertemente de la fracción de la población incluida en la muestra. Al utilizar la corrección para poblaciones finitas $(1 - n/N)$, se está tomando en cuenta la fracción de población muestreada, y por ello podría pensarse que es relevante el cálculo del tamaño de una muestra de esta forma.

Cuando se muestrea el 10% o menos de una población, el ajuste puede tener un efecto favorable aparente en las estimaciones de la varianza, al reducir la varianza. Sin embargo, gran parte de las encuestas por muestreo involucran fracciones muy pequeñas de la población. En tales circunstancias, incrementos pequeños en la fracción de población incluida en la muestra no tienen efecto en la mejora de las estimaciones para la generalización de los resultados de la muestra a la población, esto es, se puede ignorar la reducción de la varianza.

Cuando N es grande, el tamaño de la población de donde se obtiene una muestra de un tamaño en particular, por sí solo, virtualmente no tiene impacto en qué tan bien la muestra describe a la población.

Además, el impacto de la fracción de la población en muestra resulta trivial comparado con el tamaño total de la muestra y otras características del diseño, por lo que su consideración no es básica para decidir el tamaño de una muestra.

La segunda proposición equivocada es pensar, en abstracto, que el tamaño de muestra elegido para un estudio en particular, en determinada población, es útil para otro estudio diferente en el mismo lugar.

Aunque a veces ayuda ver qué tamaño de muestra consideran apropiado otros investigadores en una población en particular, la decisión del tamaño de muestra debe tomarse caso por caso, teniendo en cuenta las características propias de la encuesta en cuanto a lo que busca medirse y el tipo de estimadores y conclusiones que se espera generar a partir de cada encuesta.

Así, para determinar el tamaño de una muestra se deben tener presentes los fines de la investigación y establecer la característica o características a estimar, el grado de confianza y el error que se está dispuesto a tolerar en las estimaciones.

A todo esto hay que agregar el costo involucrado.

Si N es grande en comparación a n , la ecuación para el cálculo de n para estimar proporciones en una muestra seleccionada con un muestreo aleatorio simple es:

$$n_0 = \frac{t^2 pq}{d^2},$$

donde

d es la precisión deseada y

t es el desvío normal correspondiente al nivel de confianza fijado

Pero si N no es grande en comparación a n se ocupa

$$n = \frac{n_0}{1 + n_0/N}.$$

A continuación se explican cada uno de los conceptos.

La precisión d que se requiere en los resultados de la encuesta puede establecerse al definir la cantidad de error que se está dispuesto a tolerar en las estimaciones muestrales. El qué tanto puede variar una estimación del valor verdadero de la población (valor desconocido) y todavía ser aceptable, depende del objetivo particular de la investigación.

La precisión se mide ya sea en términos de error absoluto o de error relativo. El primero explica qué tanto se puede alejar la estimación del valor verdadero de la población,

independientemente de cuál sea éste, mientras que el error relativo toma en cuenta el valor en la población de la característica de interés, por eso el nombre de relativo. En otras palabras, si se quiere una precisión con un error absoluto de 0.1 con una $P=0.5$, entonces se dice que se acepta que los valores de la estimación de P varíen de $0.5-0.1 < p < 0.5+0.1$, esto es, $0.4 < p < 0.6$; en cambio si lo que se quiere es un error relativo de 0.1, se acepta que $0.5(1-0.1) < p < 0.5(1+0.1)$, es decir, $0.45 < p < 0.55$. El error relativo mide la precisión relativa de p como un estimador de P y no la precisión absoluta de P .

Al error absoluto se le denomina error estándar, y la fórmula para medirlo es simplemente la raíz cuadrada de la varianza del estimador. En el caso de proporciones, omitiendo la *cpf*, queda

$$s(p) = \sqrt{v(p)} = \sqrt{\frac{pq}{n-1}}$$

Para medir el error relativo, también llamado coeficiente de variación, se utiliza la fórmula

$$r = \frac{s(p)}{p} = \sqrt{\frac{q}{(n-1)p}}$$

Por su parte, el nivel de confianza, tiene que ver con la estimación por intervalos, es decir, proporciona la probabilidad de que el verdadero parámetro se encuentre dentro de cierto rango alrededor de la estimación. Es decir, de 100 muestras que se saquen en cuántas de ellas el intervalo de confianza contiene el valor poblacional.

El intervalo de confianza se calcula a través de

$$p \pm ts(p).$$

t es una función creciente del nivel de confianza. Por lo tanto, mientras más cercano al 1 (100%) se escoja el nivel de confianza se deben esperar intervalos de confianza más anchos.

Ahora, dado que estudios en la teoría de probabilidades sobre la distribución de medias de muestras aleatorias, prueban que para cualquier población que tiene una desviación estándar finita, la distribución de la media muestral (los valores que la media de la muestra puede tomar junto con su probabilidad de materializarse) tiende a la distribución normal conforme n aumenta, se puede buscar en la tabla de los valores del área bajo la curva de la distribución normal, un valor de t que represente el valor del desvío normal correspondiente al nivel de confianza deseado. Para el caso particular del 95% de confianza, $t = 1.96$.

En otras palabras, el hecho de usar las desviaciones normales t para construir el intervalo de confianza implica la suposición de que la distribución normal es una buena aproximación para la distribución de muestreo de la estadística \bar{y} .

Otra forma común para asociar un nivel de confianza la establece el siguiente teorema:

Teorema de Chebyshev⁷

Sea Y una variable aleatoria con una media finita μ y una varianza σ^2 . Entonces, para cualquier constante positiva k ,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Por ejemplo, la probabilidad de que μ esté en el intervalo $Y \pm k\sigma$, para $k = 2$, es de al menos 0.75. Aunque, usualmente la probabilidad real excede el límite inferior obtenido a partir de este teorema.⁸

⁷ Mendenhall, William. *et al. Estadística Matemática con Aplicaciones*. Grupo Editorial Iberoamérica, México, 1986. p. 118

⁸ Se ocupa la k en este teorema en lugar de la t , ya que esta es la notación de la bibliografía consultada.

Cabe mencionar también que este resultado se puede aplicar para cualquier distribución de probabilidad.

Regresando al cálculo de n , hasta ahora se tiene cómo obtener los valores de t y d . Sin embargo, el valor de la varianza de la estimación no se conoce, ya que precisamente forma parte de las estimaciones que se quieren obtener a partir de la recolección de los datos de la encuesta.

Por ello, se debe recurrir a caminos alternativos para estimar la varianza de la población a fin de determinar el tamaño de muestra. Estas opciones son:

- Ocupar los datos de un muestreo previo de la misma población o semejante
- Conjeturar respecto a la estructura de la población
- Realizar una encuesta piloto⁹
- Realizar una muestra de dos pasos¹⁰

1.4. ESTIMACION DE PROPORCIONES A PARTIR DE DIFERENTES DISEÑOS DE MUESTREO

Como se mencionó con anterioridad, además del muestreo aleatorio simple, existen diferentes diseños de muestreo que constituyen diseños más prácticos, económicos o precisos.

Estos tipos diferentes de diseño de muestras pueden combinarse, así como también llevarse a cabo en varias etapas. Esto último consiste en una jerarquía de diferentes tipos de unidades

⁹ Una encuesta piloto es aquella encuesta que se realiza para "probar el cuestionario y los métodos de campo en pequeña escala". Además, se pueden realizar estimaciones que sirvan para la planeación de la encuesta. Kish, Leslie. *Op. cit.* p. 27

¹⁰ El muestreo de dos pasos se refiere primero, a obtener una muestra de tamaño n_1 , y realizar estimaciones de una proporción o de la varianza y en base a éstas calcular el tamaño de n . Entonces el segundo paso es levantar una muestra de $(n - n_1)$ para completar el tamaño de muestra. Este método proporciona estimaciones muy confiables, pero retarda la recepción de la información de la encuesta.

de muestreo con la ventaja de que los marcos de etapas posteriores se requieren solamente para las unidades seleccionadas de etapas anteriores. En el muestreo polietápico las unidades en cada etapa se pueden elegir con un procedimiento diferente, dando lugar a una gran variedad de diseños posibles.

Sin embargo, en este capítulo sólo se presentan cuatro tipos de diseños de muestreo que se consideran básicos: muestreo aleatorio estratificado, muestreo sistemático, muestreo por conglomerados y muestreo con probabilidad proporcional al tamaño.

Para cada uno de estos diseños a continuación se describen los procedimientos de selección¹¹.

1. Muestreo aleatorio estratificado

Generalmente, poco se conoce acerca de las características de cada uno de los miembros individuales de la población antes de la recolección de datos. No es raro, sin embargo, identificar unas cuantas características de una población en el momento de muestrear. Cuando este es el caso, se separan todos y cada uno de los elementos de la población en subpoblaciones o grupos que no presenten traslapes, llamados estratos, y posteriormente se selecciona una muestra aleatoria simple independiente en cada estrato¹². De este modo se puede producir una muestra que es más probable que refleje el total de la población que un muestreo aleatorio simple.

Para el cálculo de estimaciones en un muestreo aleatorio estratificado, se incluye la ponderación W_h del estrato. El ponderador asigna a cada estrato el peso que éste tiene en la población.

¹¹ Las estimadores de proporciones y varianzas para cada uno de los tipos de diseños básicos que se tratan en esta sección se presentan en el Anexo A.

¹² La muestra obtenida en cada estrato puede realizarse con otros diseños de muestreo además que con un muestreo aleatorio simple.

Para el mismo tamaño de muestra, este procedimiento de selección usualmente produce un estimador que posee una varianza más pequeña que la que se puede obtener por muestreo aleatorio simple, si las mediciones en estudio son homogéneas dentro de los estratos, pero las medias por estrato varían entre ellas. Aunque cabe mencionar que la estratificación sólo aumenta la precisión de las estimaciones de las variables que se relacionan con las variables de estratificación.

2. Muestreo sistemático

Aunque el muestreo aleatorio simple es fácil de comprender, en la práctica se ocupa poco, a menos que la lista de elementos de la población de estudio sea corta o que ya esté numerada. No obstante, con la mayoría de las listas puede realizarse una selección mecánicamente más fácil a la que se le llama sistemática. El muestreo sistemático¹³ consiste entonces en dividir la lista de la población en n zonas de igual tamaño $k = (N/n)$ y posteriormente seleccionar en cada zona el r -ésimo elemento. Al número r se le llama arranque aleatorio y es un número entre 1 y k elegido aleatoriamente¹⁴.

Este procedimiento se ocupa en el trabajo de campo para hacer más práctica la selección de los elementos en muestra, y como consecuencia se reducen los costos en la recolección de los datos.

En este tipo de muestreo definitivamente es importante examinar el marco de muestreo potencial, con el fin de ver si hay o no alguna razón para pensar que la muestra resultante de un arranque aleatorio será sistemáticamente diferente de la generada con otro arranque, de tal forma que afectara los resultados de la encuesta. Dicho de otra manera, se debe verificar que el orden del marco muestral no esté relacionado con la variable a estimar.

¹³ Algunos autores consideran que el muestreo sistemático es solamente un proceso de selección, ya que "en un sentido estricto, una selección sistemática de unidades no es medible, porque la varianza no puede calcularse a partir de la muestra solamente." Kish, Leslie. *Op. cit.* p. 148

¹⁴ Cuando k no es entero se puede escoger entre permitir que el tamaño de muestra sea n ó $n+1$, eliminar con un método de selección con igual probabilidad suficientes unidades para reducir los listados a nk unidades, considerar la lista como circular o utilizar intervalos fraccionarios. Ver: Kish, Leslie. *Op. cit.* p. 145-147

3. Muestreo por conglomerados

En el muestreo por conglomerados la unidad de muestreo es una colección o conjunto de elementos llamado conglomerado. Generalmente se emplea por razones de costo o porque no se dispone de un marco adecuado para los elementos.

Este muestreo se puede realizar en una o en varias etapas. En el primer caso, se seleccionan conglomerados y se obtiene la medición de cada uno de los elementos que pertenecen a los conglomerados seleccionados, es decir, se hace un censo en el conglomerado seleccionado. En el segundo caso, una vez hecha la selección de conglomerados, se realiza una subselección de elementos en cada conglomerado seleccionado.

Además, para su estudio, el muestreo por conglomerados se divide en muestreo por conglomerados de tamaño igual o desigual. Cabe señalar que,

una muestra diseñada e iniciada con conglomerados iguales puede a menudo terminar con conglomerados desiguales. La desigualdad puede existir en los tamaños reales de los conglomerados que se diseñaron con medidas de tamaños iguales. También las no respuestas ocasionan desigualdades en los resultados finales del conglomerado.¹⁵

La selección de conglomerados desiguales ocasiona varios problemas. En primer lugar, el tamaño de la muestra no es fijo; se convierte en una variable aleatoria, que depende de la selección al azar de conglomerados mayores o menores, principalmente cuando el muestreo se realiza en una etapa.

Dada la forma como se calcula una proporción en muestreo por conglomerados, en donde a_i es el total de elementos en el i -ésimo conglomerado, y si se considera a éste una variable aleatoria, ya que depende de los conglomerados que resulten seleccionados, entonces para el caso de conglomerados desiguales la proporción p toma la forma de un estimador de razón:

¹⁵ *Ibidem*, p.220

$$r = \frac{\sum y_i}{\sum x_i}$$

donde

y_i es la característica de interés

x_i es la variable de conteo¹⁶

y ambas son variables aleatorias.

El muestreo por conglomerados puede ser mejor que el muestreo aleatorio simple o que el estratificado si las mediciones dentro de los conglomerados son heterogéneas y las medias entre conglomerados son aproximadamente iguales.

4. Muestreo con Probabilidad Proporcional al Tamaño (PPT)

El tamaño de muestra total puede estar sujeto a una variación mayor de la debida si se basa en una selección aleatoria de conglomerados que difieren mucho en su tamaño.

"El muestreo por conglomerados suele proporcionar una situación ideal para el uso de muestreo con PPT, ya que el número de elementos en un conglomerado, representa una medida natural del tamaño de conglomerado"¹⁷ y precisamente, como su nombre lo indica, en el muestreo con PPT cada unidad de muestreo se selecciona con probabilidad proporcional a su tamaño.

El tamaño total de muestra depende de los conglomerados que resulten seleccionados para la muestra. Por lo tanto, se debe tratar de evitar una muestra aleatoria no controlada de conglomerados con grandes variaciones en tamaño.

¹⁶ La suma de x_i equivale al tamaño de muestra n .

¹⁷ Schaffer, Richard L., et al. *Elementos de Muestreo*. Grupo Editorial Iberoamérica. México, 1987. p.217

En la mayoría de las situaciones, es innecesario e imposible tener un control exacto del tamaño de la muestra... Algunas razones son: puede resultar demasiado difícil obtener la información o los procedimientos para controlar con firmeza el tamaño de muestra inicial, el costo de recolección de datos puede imponer un límite superior en el tamaño total de la muestra y, además, las diferencias y fluctuaciones grandes en los tamaños de los conglomerados pueden causar ineficiencia administrativa en el trabajo de campo.¹⁸

Entonces, cuando las unidades difieren mucho en tamaño para tener un control del tamaño de muestra se recomienda hacer una selección con probabilidad proporcional a su tamaño. Generalmente, el muestreo con PPT solamente se utiliza en la primera etapa en un muestreo de dos etapas debido a que los elementos dentro de conglomerados tienden a ser algo similares en tamaño.

La reducción proporcional en la varianza, usando un estimador de la media para un muestreo con PPT, sería mayor si las mediciones poblacionales tuvieran más variabilidad y si las probabilidades de que los elementos aparezcan en la muestra estuviesen más próximas a ser verdaderamente proporcionales al tamaño de las mediciones poblacionales.

1.5. EFECTO DE DISEÑO

Después de haber obtenido una muestra con un diseño de muestreo complejo, es útil calcular el efecto del diseño (*ef_d*), es decir, el cociente de la varianza de la estimación del procedimiento de selección utilizado y la varianza de la estimación con una selección aleatoria simple para un mismo tamaño de muestra.

En realidad, el efecto de diseño se calcula para algunas características importantes de la encuesta, pero si se puede suponer que el efecto de diseño es constante en todas las

¹⁸ Cf. Kish, Leslie. *Op. cit.* p. 259

estadísticas de la encuesta, se facilita el cálculo del tamaño de muestra con el diseño complejo y se puede ver si es preferible en cuanto al costo a un muestreo aleatorio simple.

Un valor mayor que uno en el cociente, significa que el diseño de muestreo utilizado provoca una mayor varianza que si se empleara un muestreo aleatorio simple. Así, un efecto de diseño de 2 dice que, para obtener la misma precisión que en una muestra con un diseño aleatorio simple, se debe levantar el doble de muestra.

Sin embargo, es difícil anticipar los efectos de las características del diseño en la precisión de las estimaciones; éstos difieren de estudio a estudio y para diferentes variables en una misma encuesta. Pero este cálculo puede servir para la planeación de otra encuesta que tenga objetivos similares y el mismo diseño.

Independientemente del diseño de muestreo que se utilice, se debe tener presente que un buen diseño de muestreo requiere que se equilibren sensatamente ciertos criterios. En particular, Kish señala cuatro¹⁹:

1. Tener presentes los objetivos de la encuesta para que con base en ellos se lleve a cabo un proceso de selección conveniente y se realicen las estimaciones necesarias.
2. Calcular estimaciones de su variabilidad de muestreo, expresada con los errores estándar, estimación que, como se mencionó antes, es la base necesaria para las inferencias estadísticas.
3. Cuidar que el diseño sea práctico y capaz de llevarse a cabo en el operativo de campo, considerando la construcción y uso de marcos de selección o listas de la población.
4. Buscar el cumplimiento de los objetivos de la encuesta con un costo mínimo.

¹⁹ *Op. Kish, Leslie. Op. cit. p. 45-46*

Los cuatro criterios suelen estar en conflicto, y el experto en muestreo debe equilibrarlos y mezclarlos para obtener un buen diseño de muestreo. Sin embargo, no existe una definición única de una muestra buena o deseable.

1.6. PRECISION EN LA ESTIMACION DE PROPORCIONES

En la sección 1.3 se mencionaron los conceptos de precisión, error absoluto y error relativo. En esta sección se discutirán algunos otros puntos relacionados con la precisión en la estimación de proporciones.

Cuando se especifica el nivel de precisión de una encuesta por muestreo, generalmente no se contempla el hecho de que habrá errores de otras fuentes diferentes a las del muestreo.

En tales casos, el cálculo de la precisión basado en errores de muestreo es una manera simplificada, pero no real en la medición de los errores, ya que existen dos tipos de errores: de muestreo y no de muestreo.

- Errores de muestreo son en los que se incurre por el hecho de ocupar una muestra y no el total de la población.

El error de muestreo es un error de no observación, y las estimaciones de las encuestas están sujetas a errores de muestreo, porque no se miden todos los miembros de la población; de ser así este error se elimina.

El estadístico usado más frecuentemente para describir estos errores se llama error estándar (de la media), y es la desviación estándar de la distribución del estimador de la media muestral que se formaría si se obtuvieran todas la muestras posibles de un tamaño dado. La

variación surge de obtener estimaciones diferentes de características de la población con grupos diferentes de elementos en la muestra, y cada observación en la distribución de muestreo es el resultado de una muestra de un diseño dado.

El tamaño y diseño de una muestra probabilística, junto con una distribución de lo que se está midiendo, determina los errores de muestreo. El error de muestreo está bajo control del diseño de la muestra; en contraste con otros errores, el diseño de muestreo puede determinar un nivel específico de error de muestreo.

- Errores no de muestreo son los que están relacionados con el instrumento de medición, la no respuesta, detalles del trabajo de campo, etc.

No obstante todos los cuidados que se ponen en las diferentes etapas que se llevan a cabo en el desarrollo de una encuesta, siempre se tienen errores no de muestreo. Sin embargo, no es posible cuantificarlos ni controlarlos con el aumento del tamaño de la muestra; por lo tanto, deben de tomarse medidas preventivas para minimizarlos.

Como se mencionó anteriormente, la varianza de la distribución muestral se conceptúa como la variación entre los valores de un estimador de la encuesta, dado un diseño de muestreo. Esto es, si cada encuesta está sujeta a errores no de muestreo similares, se mide la variación alrededor de las estimaciones, cada una sujeta a este tipo de errores.

La varianza de la distribución muestral no está en función de desviaciones del valor de la población sino de la estimación obtenida por muestreo. Luego, los intervalos de confianza construidos con base en los resultados de la encuesta (a partir de la varianza de la muestra) no reflejan ningún error no de muestreo, siendo que éstos podrían encontrarse de forma significativa en la encuesta.

El modelo más simple de una encuesta por muestreo se basa en el supuesto de que la única fuente de variación en los resultados de la encuesta proviene de medir diferentes subgrupos de la población. Así la varianza de muestreo es la única variable de error.²⁰

Sostiene que si la encuesta se repitiera en la misma muestra de elementos, se obtendría el mismo resultado, considerando los errores no de muestreo constantes sobre todas las muestras posibles, dado un diseño de muestreo.

En este planteamiento se concentra la atención en un error de la encuesta, la varianza de la muestra, para la que la teoría de muestreo ha desarrollado técnicas para su medición. Sin embargo, es importante mencionar que no debe centrarse la atención en el cálculo del error de muestreo. No deben descuidarse el desarrollo del cuestionario, el entrenamiento a los entrevistadores, los errores no de muestreo en general.

En la sección 1.3 también se dijo que la precisión se establece de acuerdo a los objetivos de la encuesta. Se podría pensar que entre mayor precisión mejor; sin embargo, esto no necesariamente es válido si se toma en cuenta que al aumentar la precisión entonces también se incrementa el valor de n_0 .

En la ecuación en que se relacionan n (cuando N es grande en comparación a n) y la precisión deseada

$$n_0 = \frac{t^2 pq}{d^2},$$

se puede ver que al aumentar la precisión se reduce el valor del denominador, ya que éste representa, como se había visto, qué tanto se acepta que se aleje la estimación del valor real, y por lo tanto, el dividir un número fijo por un número más chico aumenta el valor de n_0 .

²⁰ Cf. Blalock, Hubert M. *Social Statistics*. 2 ed. McGraw-Hill. New York, 1979. p. 13

Así, se concluye que si bien es cierto que se debe cuidar la precisión de las estimaciones, no se debe pedir más precisión de la que en realidad se necesita porque se puede incurrir en un incremento notable de los costos de la encuesta.

En otras palabras, una muestra será demasiado pequeña si sus resultados carecen de la precisión suficiente para contribuir de manera apreciable en las decisiones. Una muestra será demasiado grande si sus resultados son más precisos de lo que se requiere para sus usos probables, o también si los errores no de muestreo sobrepasan abundantemente la precisión de la muestra.

2. REVISION EMPIRICA DE FACTORES QUE INFLUYEN EN LA PRECISION DE ESTIMACIONES DE PROPORCIONES

Así como en otras áreas, en muestreo los conceptos no son independientes sino que están interrelacionados. Azorín lo ejemplifica bien diciendo que el diseño óptimo de la muestra, en particular la determinación previa de su tamaño óptimo, sólo podría conseguirse a partir del conocimiento de la población.

En diversos libros se habla de que la precisión depende, además del diseño de muestreo, únicamente del tamaño de la muestra y no del tamaño de la población, como se verá en la siguiente sección. Sin embargo, en este capítulo se incluye también a la varianza y presencia de la característica de interés, ya que propiamente la precisión se da en términos de error absoluto o relativo, y éstos a su vez involucran a la desviación estándar.

En el capítulo anterior se trataron conceptos como la corrección para poblaciones finitas, los errores de muestreo, la distribución normal y la varianza. En este capítulo se van a retomar y se tratarán de hacer algunas aclaraciones de su intervención o su relación con la precisión.

2.1. TAMAÑO DE MUESTRA

A continuación se presenta un ejercicio realizado por Kish en donde se señala que si se mantienen constantes s^2 y la precisión deseada d^2 , el tamaño de la muestra n , también queda fijo.

$$n_0 = \frac{t^2 \frac{s^2}{Nf}}{d^2}$$

Entonces la fracción de muestreo f varía inversamente con el tamaño de la población, puesto que

$$f = \frac{t^2 s^2}{d^2 N n_0}$$

y ya que

$$n = \frac{n_0}{1 + n_0/N},$$

se deduce entonces que n varía en función de N .

Con esto se pretende explicar que el tamaño de la población afecta solamente al factor $(1 - n/N)$, del que generalmente se hace caso omiso al diseñar una muestra, pero que el uso de la cpf provoca de manera indirecta que el tamaño de la muestra esté en función de N . Así, cuando el efecto del factor de corrección para poblaciones finitas es significativo, la precisión de una estimación por muestreo se determina no sólo por el tamaño absoluto n de la muestra, sino también por la proporción de la población que se incluye en la muestra.

Sin embargo, es importante resaltar que en la mayoría de los casos la precisión depende (además del diseño de muestreo) del tamaño de la muestra, pues generalmente n es muy pequeña comparada con N con lo que f es muy pequeña y, por lo tanto, la influencia del tamaño de la población en el tamaño de la muestra no es relevante.

Una vez establecida la relación que guardan el tamaño de muestra y el tamaño de la población, se expone la relación entre el tamaño de muestra y la precisión.

A partir de

$$n_0 = \frac{t^2 pq}{d^2}$$

obsérvese que, aproximadamente (prescindiendo de la corrección para poblaciones finitas), n es inversamente proporcional al cuadrado de la precisión y, por tanto, para reducir el error a la mitad hace falta una muestra no doble, sino cuatro veces mayor.

Como se puede ver en el cuadro 2.1, para $p = 0.5$ y $n = 50$, se tiene un error absoluto de 0.14. Para conseguir un error absoluto de 0.07 con la misma p se debe tener una $n = 200$ aproximadamente.

Cuadro 2.1
Error de muestreo en relación al tamaño de muestra y la proporción

n	p								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
35	0,101	0,134	0,154	0,165	0,168	0,165	0,154	0,134	0,101
50	0,084	0,112	0,128	0,137	0,140	0,137	0,128	0,112	0,084
75	0,068	0,091	0,104	0,112	0,114	0,112	0,104	0,091	0,068
100	0,059	0,079	0,090	0,097	0,098	0,097	0,090	0,079	0,059
200	0,042	0,056	0,064	0,068	0,069	0,068	0,064	0,056	0,042
300	0,034	0,045	0,052	0,056	0,057	0,056	0,052	0,045	0,034
500	0,026	0,035	0,040	0,043	0,044	0,043	0,040	0,035	0,026
1000	0,019	0,025	0,028	0,030	0,031	0,030	0,028	0,025	0,019
1500	0,015	0,020	0,023	0,025	0,025	0,025	0,023	0,020	0,015

* La variabilidad reflejada en la tabla describe el error potencial que se tendría en las estimaciones por el hecho de utilizar una muestra con una nivel de 95% de confianza. Los cálculos no incluyen estimaciones del error de cualquier otro aspecto del proceso de una encuesta. La tabla y las ecuaciones en las que se basa se aplican a muestras obtenidas con un diseño aleatorio simple.

Cuando un porcentaje pequeño de un universo finito se incluye en la muestra, digamos 5%, la reducción de la varianza debido a la *cpf* se puede ignorar. Sin embargo, el efecto de la

cpf es significativo en los casos donde 25-50% o más de la población se incluye en la muestra.

Existe una relación importante entre precisión y tamaño de muestra ya que, al disminuir el error de muestreo, aumenta el tamaño de la muestra, luego se puede alcanzar cualquier precisión deseada con una muestra suficientemente grande.

Así, las muestras más grandes generan intervalos de confianza más cortos para las estimaciones, pues disminuye el error de muestreo, es decir, aumenta la precisión de la estimación. Por ello, "cuanto más cortos se quieran los intervalos de confianza, más grande se debe escoger el tamaño n de la muestra."¹ Sin embargo, añadir un número dado de casos a una muestra reduce el error de muestreo, más cuando la muestra es pequeña que cuando es comparativamente grande, si se toma en cuenta que al aumentar 50 casos a una muestra de 50 se está doblando el tamaño de la muestra y, por lo tanto, el cambio en el error es mayor que si se aumentan 100 casos a una muestra de 400, en donde el incremento de la muestra es del 25%.

Hay dos razones para que las muestras probabilísticas con diseños de muestreo complejos sean típicamente grandes. Primera, la organización y las operaciones necesarias para obtener muestras complejas de poblaciones grandes son tan pesadas y costosas que solamente muestras grandes pueden justificar económicamente su utilización. Segunda, la mayoría de las muestras para encuestas complejas tienen como meta medir efectos que son demasiado pequeños para poderlos detectar con muestras pequeñas.²

Las muestras probabilísticas complejas son esencialmente instrumentos para las encuestas grandes. Las muestras pequeñas de poblaciones pequeñas pueden seleccionarse con muestreo aleatorio simple, y entonces se puede utilizar la vasta teoría de que se dispone.

¹ Kreyzig, Erwin. *Introducción a la Estadística Matemática. Principios y Métodos*. Limusa, México, 1987, p. 197

² Kish, Leslie. *Op. cit.* p. 37

En muchas situaciones prácticas, la distribución de muestreo de la media estimada y de la mayoría de los estimadores comúnmente utilizados para presentar resultados de una encuesta es aproximadamente normal.

Esta tendencia a la normal de la distribución de muestreo cuando las muestras son de tamaño grande no se basa en la normalidad de la distribución de los elementos en la población. Por el contrario, las distribuciones de las características de la encuesta en la población están generalmente muy lejos de la normal. No obstante, aun en muestras pequeñas, la aproximación es buena para muchas variables comunes. A menudo las distribuciones de muestreo de las medias de muestras aun de tamaño moderado, se aproximarán suficientemente a la normal como para poder hacer inferencias estadísticas, a partir de estimaciones de la muestra, a valores de la población, y esto es consecuencia directa del Teorema Central del Límite.³

Teorema Central del Límite⁴

Si se obtienen repetidas muestras de tamaño n de cualquier población (de cualquier forma) con una media μ y una varianza σ^2 , entonces tanto como n crezca, la distribución de muestreo de la media de la muestra se aproxima a la normalidad, con una media μ y una varianza σ^2/n .

La aproximación a la normalidad de la distribución de muestreo de una estadística es más rápida para unas variables que para otras. Una aproximación más rápida a la normalidad indica que hay mejor aproximación de la distribución de muestreo para un tamaño de muestra dado, o que la aproximación deseada se puede alcanzar con un tamaño de muestra menor.

Es importante resaltar que además de la precisión existen varios factores que, directa o indirectamente, afectan la determinación del tamaño de muestra y que se deben tomar en cuenta:

³ Ibidem, p. 34

⁴ Blalock, Hubert M. *Op. cit.* p. 183

1. Recursos disponibles

El tamaño de muestra debe ser necesariamente consistente con los recursos disponibles para realizar la investigación. Para ello se requiere una estimación del costo, tiempo y materiales que serán necesarios para un determinado tamaño de muestra.

2. Número de características que se investigan

Dados el costo y el trabajo administrativo y de campo que se involucran en el desarrollo de una encuesta, en la mayoría de las encuestas se busca captar información de más de una característica.

Una forma de determinar el tamaño de muestra es considerar márgenes de error para cada una de las características captadas más importantes, y escoger el valor más grande de n si éste no sobrepasa el presupuesto. Sin embargo, cuando el número de variables es grande y se especifica un grado de precisión para cada una, los cálculos conducen a una serie de tamaños de muestra diferentes y por lo tanto conflictivos; una opción es disminuir la precisión en algunas características para así reducir n .

3. Estimaciones para subpoblaciones

Al determinar el tamaño de la muestra se debe tomar en cuenta el número de categorías o clases en las que se van a agrupar o analizar los datos. El tamaño de muestra total escogida debe ser lo suficientemente grande para asegurar tamaños de muestra que representen bien a las subclases independientes más pequeñas.

4. Homogeneidad o heterogeneidad de la población

Entre más homogénea sea una población se requerirá una muestra más pequeña; por el contrario, entre más heterogénea sea una población el tamaño de muestra deberá ser mayor.

Los casos de estimaciones para subpoblaciones y homogeneidad o heterogeneidad de la población se tratan en las siguientes secciones de este capítulo.

2.2. TAMAÑO DE MUESTRA PARA SUBPOBLACIONES

Hasta ahora se ha dicho que para estimar una proporción se separan los elementos en dos grupos: los que poseen y los que no poseen determinada característica de interés. Además se dijo que α representa la suma de todos los elementos que poseían dicha característica. Esto es lo que se conoce como clasificación en dos clases. Sin embargo, en la práctica es común clasificar los elementos en más de dos clases. Entonces se diría que a_i es el grupo de elementos en la muestra que tienen el i -ésimo atributo, y

$$p_i = \frac{a_i}{n}$$

es la proporción de elementos en la muestra que tienen el i -ésimo atributo.

Si lo que se quiere es una encuesta para proporcionar información con precisión estadística propia para alguna subpoblación, a esa subpoblación se le llama dominio de estudio.

Así, la subpoblación es una parte de la muestra, que se considera como muestra de la porción correspondiente de la población de la encuesta. Como tal, pueden considerarse estimaciones de la varianza y precisión diferentes en cada dominio.

Dependiendo de la información que se considere para establecer las subpoblaciones en la muestra, se puede o no saber con anterioridad a qué dominio pertenece cada elemento de la muestra.

Cochran establece que para determinar el tamaño de la muestra total, cuando se tienen varios dominios de estudio, se debe considerar lo siguiente:⁵

1. Si los dominios se pueden identificar con anterioridad, se calcula por separado la n necesaria en cada dominio. En caso de querer estimaciones con varianza y precisión iguales en cada una de las k subpoblaciones, el tamaño de la muestra debe ser k veces más grande que el necesario para un estimador global con la misma precisión (kn).

Pero, si en lugar de levantar una muestra de tamaño kn , se mantiene una muestra de tamaño n para k dominios, el error en la estimación en cada subpoblación se verá afectado por \sqrt{k} en forma directamente proporcional.

2. Si los dominios no se pueden identificar con anterioridad, pero se conocen las proporciones π_i de las unidades que pertenecen a diferentes subpoblaciones, entonces en un muestreo aleatorio simple, el tamaño de la i -ésima subpoblación es $n\pi_i$, y la varianza promedio de la proporción en dicha subpoblación, si $n\pi_i$ es grande, es

$$v(p_i) = \frac{p_i q_i}{n\pi_i}.$$

Hay estudios en los que se requiere que la varianza de las proporciones de cada par de subpoblaciones no se traslapen, es decir,

$$v(p_i - p_j) \leq V.^6$$

Para estos casos,

⁵ Cochran, William G. *Op. cit.* p.119

⁶ La V corresponde a t^2/d^2 según la notación anteriormente utilizada.

$$n \cong \frac{1}{V} \left(\frac{s_i^2}{\pi_i} + \frac{s_j^2}{\pi_j} \right)$$

donde

$$V = \frac{r^2}{f^2} \max(p_i^2, p_j^2)$$

y r es el error relativo.

2.3. VARIANZA Y PRESENCIA DE LA CARACTERÍSTICA DE INTERÉS EN LA POBLACION

La varianza de la muestra no depende solamente de su tamaño n , sino también del diseño de la muestra. La varianza de las proporciones en un muestreo aleatorio simple está dada por

$$v(p) = \frac{pq}{n-1} (1-f)$$

y como se puede ver en el cuadro 2.2 la varianza alcanza sus valores más grandes para un tamaño de muestra dado, cuando la población está igualmente dividida entre dos clases ($p=0.5$), y la distribución es simétrica respecto a este punto. La varianza cambia relativamente poco cuando p está entre 0.3 y 0.7. En otras palabras, la varianza es más pequeña tanto como las proporciones se alejen de 0.5.

Cuadro 2.2
Varianza en relación a la proporción para una n fija

	p								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
pq	0,09	0,16	0,21	0,24	0,25	0,24	0,21	0,16	0,09

Dado que para obtener la varianza de p se debe dividir p por $(n-1)$, luego si se fija el valor de pq y se incrementa el valor del tamaño de muestra, al ir dividiendo cada vez por un número mayor, se reduce el valor de la varianza. Entonces, el incremento de la muestra siempre reduce los errores de muestreo.

Si ahora, se conjuntan los resultados del cuadro 2.1 (p. 27) y el cuadro 2.2, se puede concluir que una precisión alta (un error de muestreo pequeño) se relaciona con una varianza baja; y que a mayor precisión, es decir, menor error, se incrementa el valor de n .

Dicho de otra forma, si se obtienen todas las muestras posibles de tamaño n , las estimaciones de estadísticas descriptivas (ej. media) formarán una distribución normal alrededor del valor real de la población. A mayor tamaño de muestra y menor varianza de la variable que se está midiendo, las estimaciones caerán más cerca del valor real de la población y se obtendrán mejores estimaciones basadas en una muestra.

Cuando n es un porcentaje muy alto de N , $f = n/N$ tiende a 1 y $(1-f)$ tiende a 0. Ahora, si $n = N$, es decir, si se realiza un censo en la población, $(1-f) = 0$, y al multiplicarlo por la varianza resulta que no se tiene error de muestreo. Esto es correcto ya que no se comete error de no observación al incluir todos los elementos de la población dentro de la muestra. Sin embargo, si se omite el factor de cpf , se diría que se tiene un error de muestreo de σ/\sqrt{N} aunque en realidad es 0. Con esto, se observa la importancia de incluir en los cálculos el factor de corrección para poblaciones finitas cuando la fracción de muestreo es significativa.

La heterogeneidad de la población por sí misma produce la varianza sobre muestras en los valores estimados. Si la población fuera homogénea, es decir, si todos los miembros de la población compartieran los mismos rasgos, entonces todas las muestras (de cualquier tamaño) generarían el mismo valor para el estimador de muestreo. El error de muestreo se

presenta por lo tanto porque las unidades de la muestra son diferentes unas de otras y porque sólo se mide un subgrupo de la población en una encuesta por muestreo. El estadístico de interés para un investigador variará su valor sobre las diferentes posibles muestras de elementos, dado un diseño de muestreo.

En el siguiente cuadro, se puede observar que el tamaño de la muestra es mayor alrededor de $p = 0.5$, donde hay una mayor varianza. Por eso, una estimación conservadora de n es la que se obtiene al seleccionar n con una p cercana a 0.5 en el intervalo en que se espera esté p , o en el caso extremo, calcular n considerando $p = 0.5$. Además, se observa que en una precisión determinada, n disminuye si p se aleja de 0.5.

Cuadro 2.3
Tamaño de muestra en relación a la precisión absoluta y la proporción

d	p									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	
0,01	3457	6147	8067	9220	9604	9220	8067	6147	3457	
0,02	864	1537	2017	2305	2401	2305	2017	1537	864	
0,05	138	246	323	369	384	369	323	246	138	
0,1	35	61	81	92	96	92	81	61	35	
0,2	9	15	20	23	24	23	20	15	9	

*Cálculo de n de acuerdo a un muestreo aleatorio simple con un nivel de confianza de 95%. La precisión está dada en términos de error absoluto.

La fórmula de

$$v(\bar{y}) = \frac{s^2}{n}(1-f)$$

significa que la varianza de la media muestral está en función de la variabilidad de unidades y del tamaño de la muestra. Si hay más heterogeneidad en la población (donde S^2 grande), se requerirá un mayor tamaño de muestra para alcanzar el nivel deseado de error de

muestreo. A pesar del valor de S^2 , a mayor tamaño de muestra menor es el error de muestreo. En este caso el error de muestreo es una función simple del tamaño de muestra y la heterogeneidad en la población.

Con esto resalta la importancia de estimar la varianza para determinar el tamaño de muestra, ya que una vez prefijado el error máximo admisible, que representa la precisión mínima a exigir de los resultados, y el coeficiente de seguridad o confianza, se necesita conocer además la variabilidad de la población, ya que cuanto más dispersos estén los valores de la variable en estudio, más arriesgado será utilizar una muestra chica.

Sin embargo, si ahora n se calcula con la precisión dada en términos de error relativo (cuadro 2.4), el tamaño de la muestra siempre crece cuando p disminuye en determinado error. Tal comportamiento de n es porque si se busca el mismo error relativo para una característica que predomina en la población y para una característica que se presente raramente en la población, será más fácil encontrarla en el primer caso. Así, se dice que cuando la proporción de unidades de un tipo específico es pequeña en una población, puede ser excesivamente caro obtener una muestra que dé una estimación de precisión relativa alta.

Cuadro 2.4
Tamaño de muestra en relación al error relativo y la proporción

r	p								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,01	345744	153664	89637	57624	38416	25611	16464	9604	4268
0,015	153664	68295	39839	25611	17074	11383	7317	4268	1897
0,02	86436	38416	22409	14406	9604	6403	4116	2401	1067
0,025	55319	24586	14342	9220	6147	4098	2634	1537	683
0,03	38416	17074	9960	6403	4268	2846	1829	1067	474
0,04	21609	9604	5602	3602	2401	1601	1029	600	267
0,05	13830	6147	3585	2305	1537	1024	659	384	171
0,1	3457	1537	896	576	384	256	165	96	43
0,2	864	384	224	144	96	64	41	24	11

Para un tamaño de muestra fijo, el coeficiente de variación o error relativo disminuye consistentemente conforme aumenta p .

Cuadro 2.5
Error relativo en relación a la proporción para una n fija

	p								
$\sqrt{q/p}$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
	3,0	2,0	1,5	1,2	1,0	0,8	0,7	0,5	0,3

Porque siendo la fórmula para calcular el error relativo

$$r = \frac{s(p)}{p} = \sqrt{\frac{q}{(n-1)p}}$$

el peso de la varianza se balancea con el valor de la proporción. Así, aunque $p=0.1$ y $p=0.9$ tienen la misma varianza, al dividirse un número fijo por un valor chico da como resultado un valor más grande.

Si se sustituye esta fórmula de r , en la ecuación que se utiliza para calcular el tamaño de muestra n_0 , al ser el denominador, cada vez menor, el valor de n se incrementa.

$$n_0 = \frac{t^2 pq}{\sqrt{\frac{q}{(n-1)p}}}$$

A continuación, en el cuadro 2.6 se mantiene fijo el tamaño de muestra y se calcula cuál sería el error absoluto y relativo para determinado valor de la proporción:

Cuadro 2.6
Error absoluto y relativo en relación con el tamaño de muestra y la proporción

n	p								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
35									
absoluto	0,099	0,133	0,152	0,162	0,166	0,162	0,152	0,133	0,099
relativo	0,994	0,663	0,506	0,406	0,331	0,271	0,217	0,166	0,110
50									
absoluto	0,083	0,111	0,127	0,136	0,139	0,136	0,127	0,111	0,083
relativo	0,832	0,554	0,423	0,339	0,277	0,226	0,181	0,139	0,092
100									
absoluto	0,059	0,078	0,090	0,096	0,098	0,096	0,090	0,078	0,059
relativo	0,588	0,392	0,299	0,240	0,196	0,160	0,128	0,098	0,065
250									
absoluto	0,037	0,050	0,057	0,061	0,062	0,061	0,057	0,050	0,037
relativo	0,372	0,248	0,189	0,152	0,124	0,101	0,081	0,062	0,041

Así, a mayor tamaño de muestra, menor error; es decir, mayor precisión para determinada proporción, ya sea en términos de error absoluto o relativo. El error absoluto crece a medida que se acerca a $p = 0.5$ y es simétrico respecto a ese punto, en cambio el error relativo decrece consistentemente cuando p aumenta.

3. APLICACION PRACTICA DEL MUESTREO PARA PROPORCIONES EN LA MEDICION DE PREFERENCIAS ELECTORALES

Este tercer capítulo busca aplicar la teoría de muestreo para proporciones en una encuesta que permita medir preferencias electorales.

Primero se describe cada una de las encuestas aplicables en México y en el mundo para obtener estimaciones de resultados electorales: el Acopio Inmediato de Resultados Electorales (AIRE) y la Encuesta Posvoto (EP).

En seguida se explican ambas metodologías: diseño de muestreo, operativo de campo y procesamiento de la información; de acuerdo a levantamientos de AIRE y EP realizados por la empresa mexicana Opinión Profesional, S.A. de C.V., esta empresa ha realizado a la fecha un número considerable de este tipo de encuestas, por lo que cuenta con la infraestructura y experiencia en el desarrollo de estos levantamientos.

3.1. MECANISMOS DE OBTENCION DE INFORMACION ELECTORAL

Como se mencionó anteriormente, las encuestas se ocupan en una gran variedad de disciplinas y con objetivos muy diversos. En relación al campo de las encuestas de opinión, Azorín refiere que

el sondeo o medición de la opinión pública ha adquirido gran importancia en muchos países, pero no únicamente por su relación con objetivos económicos, políticos y sociales. Los "investigadores del mercado", los "trabajadores sociales", los "doxólogos" (estudian las opiniones), "pséfólogos" (estudian

las elecciones o preferencias electorales) y "analizadores" de la opinión pública constituyen ya un grupo profesional establecido.¹

Enfocados a los objetivos políticos, se han efectuado muestreos para medir preferencias y evaluar el prestigio de personalidades y partidos, así como para análisis de votaciones y otras reacciones políticas y sociales.

En muchas partes del mundo, entre los estudios de mayor relevancia en este campo están los proyectos realizados con el fin de generar estimaciones, con precisión estadística específica, del porcentaje de votos obtenido por cada uno de los partidos contendientes en la jornada electoral el mismo día de las elecciones.

Debido al interés de las cadenas televisivas norteamericanas de generar estimaciones sobre el número de votos obtenidos por los diversos partidos, y de informarlos a su auditorio la noche misma de la elección, a finales de los años sesenta, surgieron en Estados Unidos dos tipos de levantamientos para recabar información correspondiente a determinadas unidades electorales elegidas de acuerdo con un diseño de muestreo:

- Acopio Inmediato de Resultados Electorales (AIRE) o Conteo Rápido
- Encuesta Posvoto (EP) o Encuesta a la salida de las casillas ("*Exit Poll*" en inglés)

Específicamente, "en 1967, Warren Mitofsky y un equipo de investigadores financiados por la CBS News encontraron un medio de obtener información acerca del resultado de las elecciones y los electores. En las elecciones para Gobernador en Kentucky se realizó el primer EP, la cual se combinó con los resultados de la votación real en las casillas de la muestra, es decir un AIRE, para obtener una proyección de la votación".²

¹ Azorín. *Op. cit.* p. 293

² Mercado Gasca, Lauro Ignacio. *Estudio del Votante Mexicano por medio de las Encuestas a la Salida de las Casillas*. Tesis. El Colegio de México, Centro de Estudios Internacionales. México, 1993. p. 4

En marzo de 1990 se funda en Estados Unidos el *Voters Research and Survey* (VRS), bajo la dirección de Warren Mitofsky y cuyos resultados se difundían por la CBS, la NBC, la ABC y la CNN. Esta empresa realiza la mayoría de estos levantamientos en dicho país. Actualmente el AIRE y la EP se utilizan en muchos países del mundo, entre ellos México, desde 1990.

3.1.1. ACOPIO INMEDIATO DE RESULTADOS ELECTORALES (AIRE)

El Acopio Inmediato de Resultados Electorales consiste en recabar, de una muestra de secciones previamente seleccionadas, los resultados que se publiquen a las afueras de las casillas, mismos que se transmiten por vía telefónica a un centro de acopio, donde se procesa la información para obtener estimaciones con base en la muestra seleccionada.

Los primeros AIRE en México fueron realizados por Opinión Profesional en noviembre de 1991, en las elecciones para presidentes municipales en cinco municipios de Nuevo León.

La recopilación empieza poco después del cierre de casillas, y sólo se espera a que los funcionarios de casilla terminen de contar los votos para recabar y transmitir la información.

En el AIRE los errores no de muestreo potenciales son:

1. Coteo erróneo de los votos por parte de las autoridades de casilla
2. Copia errónea de los datos
3. Captura errónea de los datos

El error 1 no afectaría las estimaciones muestrales, si los mismos resultados se utilizan en la contabilización oficial de votos. Y los errores no de muestreo 2 y 3, con una capacitación

adecuada al personal de recolección y de captura, pueden reducirse lo suficiente como para no influir en las estimaciones.

3.1.2. ENCUESTA POSVOTO (EP)

La Encuesta Posvoto consiste en entrevistar, mediante un mecanismo de selección "aleatorio", a los votantes afuera de las casillas o secciones electorales seleccionadas en la muestra, una vez que ya han emitido su voto. Las preguntas del cuestionario giran en torno a características socioeconómicas básicas, preferencia electoral y razón de la preferencia, así como su opinión sobre diversas situaciones, a diferencia del AIRE en donde se capta únicamente la preferencia electoral.

Las EP pueden tener dos objetivos principales:

1. Estimar el resultado de una elección
2. Conocer algunas características y opiniones de los votantes sobre algunos temas.

La encuesta se realiza a la salida de las casillas el día de las elecciones, una vez que los ciudadanos han emitido su voto (para no influir en él), porque se considera un momento muy oportuno para captar la información que se necesita para cumplir los objetivos de una EP. Primero, porque el entrevistado ya decidió por qué partido votar y porque, dada la cercanía de la acción, es fácil que recuerde el partido por el que votó. Segundo, se tiene la certeza de que las personas entrevistadas son votantes sin necesidad de pedir al entrevistado ninguna clase de comprobante. Esta situación dificultaría el trabajo del encuestador. Además, bajo este esquema, se tendrían que hacer varios intentos antes de encontrar una persona que haya ido a votar el día de las elecciones para poder obtener algunas

características y opiniones de los votantes en algunos temas, lo que repercutiría en el costo y tiempo de recepción de la información.

Las primeras EP en México fueron realizadas por Opinión Profesional en el año de 1990, en seis de los municipios conurbados al Distrito Federal, durante las elecciones para presidentes municipales en el Estado de México. El objetivo fue estimar la distribución porcentual de la votación en cada municipio y darla a conocer la noche misma de la elección.

La metodología utilizada en México ha evolucionado con base en las experiencias y en el estudio de las EP realizadas en nuestro país y en otros países del mundo.

En una EP las entrevistas se desarrollan durante toda la jornada electoral. En México, la jornada electoral es de aproximadamente 10 horas. En el transcurso del levantamiento se va recibiendo información de los cuestionarios levantados, misma que se va procesando para ir viendo el desarrollo de las elecciones. Una vez capturados todos los cuestionarios que se levantaron en el día, con la información de la EP se generan estimaciones finales de los porcentajes obtenidos por cada uno de los partidos contendientes.

En la EP los errores no de muestreo potenciales son, al igual que en el AIRE

1. Registro erróneo de la información
2. Captura errónea de la información

pero además, se añaden

3. Posible influencia del encuestador
4. Limitaciones en la recolección de información propias de un cuestionario
5. No respuesta
6. Respuesta falsa del entrevistado

En otras palabras, en una EP puede haber errores de entrevista, posiblemente la entonación del encuestador al leer una pregunta influya en la respuesta del entrevistado, o quizá errores del instrumento de medición, en este caso el cuestionario; éstos se refieren a aspectos como el ordenamiento de las preguntas, la redacción de las mismas, la selección de las opciones para clasificar las respuestas, etc. Ambos errores pueden ser factores que induzcan a una no respuesta al cuestionario o a la obtención de información falsa, por lo que en este tipo de encuestas debe tenerse siempre presente el objetivo de tratar de minimizar los errores no de muestreo.

3.2. METODOLOGIA DEL ACOPIO INMEDIATO DE RESULTADOS ELECTORALES

En general, las encuestas constan de tres etapas fundamentales para su realización:

- Diseño de muestreo
- Operativo de campo
- Procesamiento de la información

En esta sección se explicarán cada una de ellas referidas al Acopio Inmediato de Resultados Electorales.

3.2.1. DISEÑO DE MUESTREO

Para poder definir el diseño de muestreo, antes se debe considerar: objetivo de la encuesta, marco de muestreo y las características y los dominios de estudio de interés. Luego, una vez especificado el método de selección de la muestra, se determina su tamaño. A continuación se enuncian cada uno de estos puntos enfocados al AIRE.

- **Objetivo de la encuesta**

El objetivo del AIRE, como se mencionó con anterioridad, es estimar la distribución del voto del electorado en una determinada jornada electoral la noche misma del día de las elecciones.

- **Marco de muestreo y características de interés**

La población objeto de estudio está constituida por las personas mayores de 18 años que acuden a las urnas a emitir su voto el día de las elecciones en el área geográfica de interés, y la característica de interés es el partido por el que votan, es decir, su preferencia electoral.

"Dado que la participación en los comicios es decisión personal de los ciudadanos, no es factible utilizar como marco de muestreo las listas de empadronados, ya que no se conoce de antemano cuáles de ellos serán votantes y cuáles no".³

Por tanto, las unidades básicas del marco de muestreo son las secciones o casillas⁴, es decir, conglomerados de votantes.

³ Opinión Profesional. *Acopia Inmediato de Resultados Electorales por Muestreo. Nota Metodológica*. México, 1993.

⁴ Una sección electoral es la unidad de máximo nivel de desagregación para la que el Instituto Federal Electoral (IFE) tiene el dato explícito de número de ciudadanos registrados. Sin embargo, una sección puede subdividirse en casillas, normalmente denominadas básicas, contiguas, especiales o extraordinarias.

• **Dominios de estudio**

En México se llevan a cabo elecciones para: Presidente de la República, Gobernadores, Senadores, Asambleístas, Presidentes Municipales y Diputados. De acuerdo al nivel de gobierno para el que se realicen las elecciones en nuestro país, se determinan los dominios de estudio, además de considerar la información que le interesa al patrocinador.

Así, el nivel geográfico de los dominios de estudio para cada nivel de gobierno es el siguiente:

Presidente de la República	Nacional
Gobernadores	Estatal
Senadores	Estatal
Asambleístas	Distrito Federal
Presidentes Municipales	Municipios
Diputados (federales y locales)	Distritos

De acuerdo a lo anterior, y considerando el costo de levantamiento, la empresa Opinión Profesional utiliza normalmente para la realización del AIRE el siguiente diseño de muestreo: estratificado, bietápico o polietápico (según el dominio de estudio requerido), de conglomerados, con probabilidad proporcional al tamaño y selección sistemática.

Estratificado porque se estratifican secciones⁵ o municipios en cada dominio de estudio, de acuerdo a su preferencia electoral obtenida en comicios realizados con anterioridad.⁶ Esto se hace con el fin de garantizar en la muestra la inclusión de secciones de las diferentes preferencias electorales observadas.

⁵ Opinión Profesional cuenta con la información en el nivel de sección sólo para las elecciones realizadas en 1991.

⁶ También puede especificarse una estratificación geográfica o una estratificación por tipo: urbana o rural.

Bietápico porque en caso de que el dominio de estudio sea un estado, generalmente primero se estratifican municipios, y en otra etapa se seleccionan secciones dentro de los municipios en muestra. Si se hiciera en una etapa, en un caso extremo podría caer una sección en cada municipio y esto elevaría notablemente el costo del levantamiento.

Polietápico porque a medida que aumenta el dominio de estudio, se consideran más etapas de selección.

De conglomerados porque la última etapa de selección la constituyen conglomerados de votantes, y al recabar la información del acumulado del total de votos de la casilla se realiza censo dentro del conglomerado, lo que equivale a lo que se había denominado muestreo por conglomerados en una etapa.

Finalmente, las secciones se seleccionan con probabilidad proporcional a su tamaño y de manera sistemática, para garantizar que todo tipo de tamaños de secciones se encuentren en la muestra. La medida del tamaño para cada sección es el padrón o lista nominal de la misma, según la información de que se disponga.

- **Tamaño de muestra**

Para determinar el tamaño de muestra requerido para un AIRE, Opinión Profesional clasifica el comportamiento de las preferencias electorales en:

Predominio del partido X	Si existe un partido que cuente aproximadamente con 60% o más de los votos
---------------------------------	--

Competencia entre partidos X y Y Cuando los votos se dividen entre dos partidos en donde la diferencia de porcentajes entre ellos no supera los 10 puntos porcentuales y la suma de ambos recauda 75 % o más de la votación total.

Pluripartidismo En este caso la distribución de votos entre las tres primeras fuerzas electorales es muy parecida y acumulan el 80 a 90% de la votación total.

El nivel de confianza utilizado en la mayoría de los proyectos ha sido del 95%. La precisión ha variado, pero lo más común han sido cuatro puntos absolutos cuando un partido predomina; en el caso de partidos competidos y pluripartidismo, la precisión relativa se establece por

$$\sqrt{(p_i - p_j)} \leq V$$

en donde se hacen intervalos para p_i y p_j con dos errores estándar

$$p_i \pm (2p_i r) \quad \text{y} \quad p_j \pm (2p_j r)$$

y el error relativo r es

$$r = \sqrt{\frac{(efd)^2 q}{np}}$$

donde efd es el efecto del diseño (1 en muestreo aleatorio simple), y se busca que ambos intervalos no se traslapen.

Con estos datos, ya se podría calcular el tamaño de muestra para cada estrato, si se tratara de un muestreo aleatorio simple, en cada casilla fuera a votar una persona y no hubiera no respuesta.

Pero siendo que no se ocupa un muestreo aleatorio simple, se involucra en la fórmula del tamaño de muestra el efecto del diseño que, de acuerdo al estudio de levantamientos anteriores, Opinión Profesional ha estimado que es de 2.02.

El valor de n obtenido se divide por 250, una estimación conservadora del número de votos por casilla, en virtud de que en la última etapa de selección se tienen conglomerados; finalmente se divide por 0.95, lo que equivale a suponer un 5% de no respuesta.⁷

Cabe mencionar que estos cálculos sirven de base en la determinación del tamaño de muestra, pero también intervienen los costos y la experiencia y juicio del investigador.

3.2.2. OPERATIVO DE CAMPO

Para poder tener un mejor desempeño por parte de los recolectores, días previos al levantamiento,

una vez seleccionadas las casillas en muestra, se procede a la localización de las mismas. El reconocimiento de las casillas se realiza primero en la cartografía disponible y después, físicamente en campo. Además, se localizan todos los medios de comunicación disponibles alrededor del área de trabajo, y se determina si es necesaria la contratación de teléfonos particulares para el día de la elección, con el fin de evitar el retraso en la transmisión de la información.⁸

⁷ En el AIRE se considera no respuesta al conjunto de votos que se clasifican como anulados o en blanco, ya que de éstos no se puede obtener la preferencia electoral de los votantes.

⁸ Opinión Profesional. *Op. cit.*

Con estas referencias y con la información de si la casilla está en un área urbana o rural, el tiempo de traslado de una casilla a otra y la facilidad de acceso a los medios de transmisión, se distribuyen cargas de trabajo al personal de recolección. Algunos criterios que se siguen son: cuando en la sección hay más de una casilla, se recolecta la información de todas y, siempre que la cercanía de las secciones en muestra lo permita, un recolector recoge la información de varias secciones.

Como se mencionó antes, con el fin de reducir los errores no de muestreo, los recolectores reciben una capacitación previa al levantamiento, y ya

el día previo a la elección, por la mañana, cada recolector debe ubicar sus casillas, y reportarse al centro de acopio para probar el operativo de comunicación. Reciben como material de campo un plano de localización con indicaciones de vías y medios de acceso. La tarde del día de la elección, cuando los funcionarios de casilla publican los resultados de la misma, el recolector recoge la información de la sección y la transmite inmediatamente al centro de acopio.⁹

En el centro de acopio se captura y procesa la información para producir la estimación de resultados.

3.2.3. PROCESAMIENTO DE LA INFORMACION

Para poder generar resultados, la información muestral recabada debe ponderarse de acuerdo al diseño de muestreo utilizado.

Por ejemplo, si se levanta una muestra de un AIRE para generar resultados al nivel estatal, generalmente los pasos para la selección de la muestra son:

⁹ *Ibid.*

1. Estratificación de municipios
2. Selección de a_h municipios con PPT dentro de cada estrato
3. Selección de b_h secciones con PPT dentro de cada municipio seleccionado

Entonces la probabilidad de selección de cada sección está dada por

$$w_{hj} = \frac{a_h M_h}{M_h} \frac{b_h M_{hj}}{M_h}$$

donde

- w probabilidad de selección
- h estrato
- i municipio
- j sección
- a número de municipios a seleccionar en el estrato
- b número de secciones a seleccionar en cada municipio
- M número de personas empadronadas

Ahora, si se sabe que el número de secciones seleccionadas en el h -ésimo estrato es

$$n_h = a_h b_h,$$

entonces

$$w_{hj} = \frac{n_h M_{hj}}{M_h}$$

Por lo tanto, el ponderador para estimar totales para cada sección dentro del estrato está dado por

$$F_{hj} = \frac{1}{w_{hj}} = \frac{M_h}{n_h M_{hj}}$$

Es importante mencionar que, si bien en este caso en particular se estiman totales en el proceso para estimar proporciones, esto no implica que la muestra permita estimar totales con la misma precisión con la que estima proporciones. En realidad, la estimación de totales tiene menor precisión que la que se fija para la estimación de proporciones.

Luego, la estimación de proporción de votos obtenidos por cada partido está dada por:

$$P_k = \frac{\sum_h \sum_i \sum_j F_{hij} V_{hij}}{\sum_h \sum_i \sum_j \sum_k F_{hij} V_{hij}}$$

donde

p	proporción de votos
h	estrato
i	municipio
j	sección
k	partido
F	ponderador
V	número de votos

Nótese que al utilizar F_{hij} ya no es necesario ponderar la información con un W_k .

F_{hij} es el ponderador planeado, es decir, el que se desprende del diseño de muestreo. Sin embargo, en el ponderador final, de acuerdo a la información que ha llegado hasta el momento de procesar, se involucran dos factores de corrección.

El primero busca reproducir la información en el nivel total de casillas que componen la sección. Como se mencionó en la sección 3.2.1., las secciones electorales pueden dividirse a

su vez en casillas. Este factor se ocupa cuando el número total de casillas es diferente al número de casillas de las que se ha recibido información.

El uso de este factor se basa en el supuesto de una distribución similar de los votantes en las diferentes casillas de una sección. Estudios realizados por Opinión Profesional permiten asumir que la distribución de los votos no difiere entre casillas básicas y contiguas, lo que permite el uso de dicho factor.

El segundo factor de corrección sirve para expandir al total del estrato los valores de la distribución del voto a partir de las secciones de las que se ha recibido información. Se utiliza siempre y cuando la información recibida no rebase los niveles de la no respuesta planeados. De ser así, no se utiliza este factor, ya que no se pretende expandir al total del estrato a partir de un número no suficiente de secciones.

Una vez realizado el procesamiento de la información recibida, Opinión Profesional genera tres tipos de resultados:

Porcentaje sobre el total de votos: considera tanto los votos para los diferentes partidos, como los votos anulados y los votos en blanco.

Porcentaje sobre votos válidos: no toma en cuenta las boletas en blanco

Porcentaje sobre votos efectivos: no considera ni las boletas en blanco, ni los votos anulados.

El último, es el que se transmite a los medios de comunicación contratantes para su difusión pues los resultados oficiales se emiten bajo este criterio, lo que permite una comparación directa entre ambos resultados.

3.3. METODOLOGIA DE LA ENCUESTA POSVOTO

La metodología de una Encuesta Posvoto es muy similar a la de un AIRE, con la diferencia de que en la EP existe una etapa más en la selección de la muestra, que es la selección de los votantes a entrevistar. Por lo tanto, más que explicar otro diseño de muestreo y otro procesamiento de la información, a continuación sólo se señalan las consideraciones adicionales que se hacen en una EP, respecto del AIRE.

3.3.1. DISEÑO DE MUESTREO

Para poder definir un diseño de muestreo, se deben considerar los objetivos de la encuesta, el marco de muestreo y las características y los dominios de estudio de interés. Por dicha razón se enuncian éstos en relación a la EP.

- **Objetivos de la encuesta**

La EP tiene dos objetivos: estimar la distribución porcentual de la votación el día de las elecciones y describir algunas características de los votantes.

- **Marco de muestreo y característica de interés**

Al igual que en el AIRE, la población objeto de estudio es el grupo de personas mayores de 18 años que acuden a votar el día de las elecciones. Por otra parte, las características de interés en esta encuesta son de tres tipos: aspectos socioeconómicos, preferencia electoral y razón de la preferencia y opinión sobre diversas situaciones.

- **Dominios de estudio**

Las contiendas electorales se realizan para elegir determinado nivel de gobierno, así que siendo un dominio de estudio una subdivisión de la población para la que se quiere

proporcionar información con precisión estadística, al igual que en un AIRE, los dominios para una EP se determinan de acuerdo al nivel de gobierno del candidato a elegir y el interés del patrocinador. De esta manera los dominios de estudio para los diferentes niveles de gobierno son:

Presidente de la República	Nacional
Gobernadores	Estatal
Senadores	Estatal
Asambleístas	Distrito Federal
Presidentes Municipales	Municipios
Diputados (federales y locales)	Distritos

Estos tres aspectos llevan a cubrir las siguientes etapas para la selección de la muestra:

1. Estratificación de secciones o municipios en cada dominio de estudio de acuerdo a su preferencia electoral en elecciones anteriores
2. Cálculo del tamaño de muestra
3. Distribución del número de secciones por estrato
4. Selección sistemática de secciones con PPT
5. Selección "aleatoria" de votantes en cada sección

Las unidades de observación en una EP son los votantes mismos, mientras que en un AIRE, la última etapa de selección son conglomerados de votantes.

- **Tamaño de muestra**

Siendo una EP un operativo muy similar en cuanto a objetivos a un AIRE, la planeación de ambos es similar; por lo tanto, la determinación del tamaño de muestra se hace bajo el mismo nivel de confianza (95%), los mismos criterios para establecer la precisión

4 puntos absolutos

$$\sqrt{(p_i - p_j) \leq V}$$

predominio de un partido

competencia entre dos partidos o

pluripartidismo

y la misma clasificación de la presencia de la característica de interés:

Predominio del partido X

Competencia entre partidos X y Y

Pluripartidismo

sólo que en lugar de considerar 250 votantes por casilla, se consideran 150, ya que en la EP se realiza un submuestreo de votantes.

Es importante mencionar que aunque en una EP se capta información adicional a la preferencia electoral, es con esta única variable que se estratifica a la población. Por lo tanto, las ganancias debidas a la estratificación estarán directamente asociadas con la mejor precisión de las estimaciones de la distribución del voto.

3.3.2. OPERATIVO DE CAMPO

Antes de mencionar las actividades del personal de campo, cabe señalar que entre las actividades de planeación de una EP, además de la selección de la muestra, se diseña el instrumento de medición: el cuestionario.

El cuestionario en su contenido ha sufrido modificaciones, pero básicamente consta de dos partes: en la primera se realiza la entrevista abarcando las características socioeconómicas y opinión sobre temas de actualidad; en la segunda se incluyen las preguntas de preferencia

electoral y razón de la preferencia. Para asegurar la confidencialidad del voto, esta segunda parte es de autollenado.¹⁰

El operativo de campo de una EP, al igual que un AIRE, consta de dos etapas: la primera de localización de casillas y la segunda es propiamente la recolección de la información.

El personal de campo lo componen encuestadores, supervisores y auxiliares. Todos reciben capacitación para la realización de su trabajo; los encuestadores recolectan la información y los supervisores y auxiliares la codifican y transmiten.

El día de las elecciones los encuestadores (uno por sección) deben ubicarse en su lugar de trabajo, antes que abran las casillas, para poder organizar su material e informar al presidente de casilla sobre su trabajo.

A las 8:00 abren oficialmente las casillas, y en ese momento el encuestador comienza su trabajo. Anteriormente se mencionó que la selección de los entrevistados se llevaba a cabo de manera "aleatoria", a continuación se describe dicho proceso: el encuestador intenta entrevistar a la primera persona que salga de la casilla después de haber votado, y mientras realiza la entrevista cuenta el número de personas que salen de la casilla. Al terminar la entrevista, el encuestador aborda a la primera persona que en ese momento salga de la casilla. Con esto se busca evitar que el encuestador escoja según su juicio a los entrevistados. Si la persona que le toca entrevistar no acepta la entrevista, entonces debe intentar con la siguiente. Este procedimiento se lleva a cabo durante toda la jornada electoral.

Los supervisores tienen a su cargo varias secciones. Cada supervisor debe recorrer dichas secciones y recoger los cuestionarios de cada uno de los encuestadores. Al término de su

¹⁰ En caso de que la persona no sepa leer, si ella lo pide, se hace por entrevista directa.

recorrido, codifica y transmite cada uno de los cuestionarios vía telefónica. Una vez terminada la transmisión de información realiza otro recorrido o "ronda".

La experiencia de Opinión Profesional en estos levantamientos muestra que aproximadamente un supervisor realiza 3 ó 4 rondas en el transcurso del día.

3.3.3. PROCESAMIENTO DE LA INFORMACION

Una vez seleccionadas las secciones de acuerdo a un diseño de muestreo de un AIRE, como no es posible entrevistar a todos los votantes se lleva a cabo un submuestreo dentro de cada sección.

En otras palabras, cuando se está en la sección, para un AIRE se capta la información de todos los elementos que pertenecen al conglomerado, y para una EP se realiza un submuestreo de elementos dentro del conglomerado, tal como se realiza un muestreo por conglomerados en una y dos etapas respectivamente.

Dado que la probabilidad de selección se conoce hasta la selección de las secciones, el ponderador para cada sección para un AIRE es válido también para una EP.

$$F_{hij} = \frac{M_h}{n_h M_{hj}}$$

donde

- F* ponderador
- M* número de personas empadronadas
- h* estrato
- i* municipio
- j* sección

Sin embargo, en el cálculo de las proporciones, no se conoce V_{hjk} , el número de personas que votaron por el k -ésimo partido en la j -ésima sección en el i -ésimo municipio ubicado en el h -ésimo estrato ya que, como se dijo anteriormente, no se puede entrevistar a todos los votantes y, en su lugar, se hace una selección aleatoria de ellos.

Al igual que en un muestreo por conglomerados de tamaños desiguales, se trata de obtener una estimación aproximada de V_{hjk} , el tamaño del conglomerado. En el procesamiento de la información se utilizan algunos factores de corrección aplicables al ponderador para lograr aproximarse a este valor.

En total son 3 los factores de corrección que se utilizan, y a continuación se describen cada uno de ellos.

Factor 1 (F_1)

Así como las secciones pueden dividirse en casillas, en las casillas pueden haber varias mesas de trabajo. El objetivo de este factor es saber qué fracción de la población empadronada de la casilla se está captando, y corregirla para que represente a la casilla en total, bajo el supuesto de una distribución uniforme del flujo de votantes en las diferentes mesas de trabajo en la casilla.

En la EP no se toma en cuenta el número de casillas que forman la sección, porque dado que generalmente se localizan en domicilios diferentes, no es posible para el encuestador entrevistar a los votantes más que de una casilla.

Factor 2 (F_2)

Antes de generar las estimaciones en cada ronda, las secciones se catalogan en lo que se denomina en sí se auto-representan o no. Se dice que una sección se auto-representa si la

información recabada en la sección contiene suficiente información como para representarla.¹¹

Cada ronda se trata de manera independiente y en cada una se identifican las secciones que no se auto-representan, estableciendo una relación entre el número de entrevistas realizadas y el "censo" o la suma de personas que salen de la casilla en el transcurso de las entrevistas.

Para cada estrato de las secciones-rondas que sí se auto-representan, se ocupa un segundo factor para corregir la ausencia de secciones en el estrato que, al igual que en un AIRE, se ocupa sólo si no se rebasa la no respuesta planeada de dicho estrato. De ser así, no se ocupa este factor, y por este concepto no se modifican los ponderadores de dicho estrato.

Para las secciones que no se auto-representan en una ronda, se busca otra ronda de esa misma sección que sí se auto-represente. En caso de que exista, se suman las entrevistas y el censo para ambas rondas (la que se auto-representa y la que no), y nuevamente se determina si se auto-representa. Si después de hacer esto, la sección sí se auto-representa, entonces se suman los votos obtenidos por cada partido. Si no se auto-representa, se toman los valores de cada ronda por separado.

Las secciones que no se auto-representan en una ronda no se involucran en el proceso, se espera la siguiente ronda de esa sección para ver si ya juntas se auto-representan. Pero si al final de la jornada electoral todavía hay secciones-ronda que no se auto-representan aun acumulando todas las rondas, únicamente en este caso, sí se involucran en el proceso.

¹¹ El criterio para esta clasificación es el siguiente: si el número de entrevistas es mayor o igual que $(\text{censo}/20)$, sí se auto-representa, en otro caso no se auto-representa.

Factor 3 (F_3)

En el transcurso de la jornada electoral, en algunas ocasiones se presentan problemas que provocan que el encuestador tenga que interrumpir su trabajo y, por ende, no pueda contar el número de personas que salen de la casilla; a esto se le llama "interrupción" en el censo. Para corregir esta falla en el censo, en las secciones en las que se hayan reportado interrupciones y si no se rebasan los niveles de no respuesta planeados, se aplica este factor de corrección.

Este factor 3 pretende estimar el número de votantes que no se contaron durante la interrupción, bajo el supuesto de que el flujo de votantes tiene una distribución uniforme en las secciones del estrato.

Finalmente, la ecuación para expandir los votos muestrales por partido a cada ronda-sección está dada por

$$V_{rhjk} = V_{rhjk} \frac{\text{censo}_r}{\text{entrevistas}_r} F_{hj}$$

donde

r	ronda
h	estrato
i	municipio
j	sección
k	partido

y al ponderador F_{hj} se le aplican los factores de corrección que sean necesarios. Es decir, que F_{hj} se llega a expresar como: $F_{hj}(F_1)(F_2)(F_3)$.

4. ANALISIS Y PRESENTACION DE RESULTADOS

El objetivo de este capítulo es realizar un análisis del comportamiento de las estimaciones de proporciones que se generan, tanto con el AIRE como con la EP, a partir de operativos recientes realizados por Opinión Profesional.

Primero se trabaja con algunos operativos de AIRE y se distinguen los diferentes niveles de presencia de la característica de interés en la población. Se retoman algunos aspectos del diseño de muestreo y, después, se comparan los valores generados por un AIRE y los resultados oficiales o poblacionales.

La segunda parte de este capítulo se enfoca a comparar las estimaciones de la información demográfica recogida en una EP y las cifras respectivas del XI Censo Nacional de Población y Vivienda. Cabe recordar que este tipo de información de la EP se transmite a los medios de comunicación para describir, la noche misma del día de las elecciones, las características de los votantes. Además, se realiza una descripción del flujo de votantes a lo largo de la jornada electoral. Finalmente, se presenta una comparación entre los resultados de AIRE, EP y resultados oficiales.

4.1. ANALISIS ESTADISTICO DE ALGUNOS OPERATIVOS DE AIRE

La presencia de la característica de interés en la población puede catalogarse como

- Predominante
- Dividida
- Rara

pero si la característica de interés es la preferencia electoral de los votantes, y se trata de un país en donde compiten más de dos fuerzas electorales, se puede añadir el rubro

- Pluripartidismo

Se trata de ejemplificar cada uno de estos casos en un AIRE con las tres primeras fuerzas electorales en México: PAN, PRI y PRD, sin ninguna referencia al orden en que se enuncian.

El caso de predominio se presenta en el levantamiento realizado en las elecciones para gobernador de Morelos, en marzo 1994, donde predomina el PRI. La votación está dividida entre dos partidos, PAN y PRI, en los comicios para presidente municipal en Mérida, Yuc., noviembre 1993; en este operativo también se ejemplifica la presencia rara de la característica de interés, es decir, la baja propensión de votos para otro partido que no sea PAN ni PRI. Y en el mismo mes, pero en el municipio de Ocoyoacac, Edo. de México, existe un pluripartidismo entre PRI, PRD y otros.¹

Los resultados obtenidos a partir del AIRE en cada uno de los operativos anteriormente mencionados son:

	MORELOS	MÉRIDA	OCOYOACAC
PAN	9.5	49.9	9.5
PRI	61.8	48.3	37.5
PRD	25.7	0.4	27.8
OTROS	3.0	1.4	25.2

¹ En "otros" se agrupan los resultados obtenidos por los otros seis partidos registrados que contendieron en la elección y los votos asignados a otros partidos no registrados. Si bien los seis partidos son diferentes y hubieran ameritado una estimación por separado cada uno, se unieron aquí para poder ejemplificar el pluripartidismo.

Ahora, de acuerdo a lo que se dijo en la sección 1.5,

$$efd = \frac{\sqrt{\text{diseño utilizado}}}{\sqrt{\text{aleatorio simple}}}$$

entonces, el estimador de la varianza de las estimaciones anteriormente citadas (omitiendo la *cpf*) es

$$v(p_k) = \frac{2.02 p_k q_k}{(n-1)}$$

donde

- k* partido
- n* total de votantes
- 2.02 efecto de diseño

luego los intervalos de confianza del 95% para dichas estimaciones son

	MORELOS	MERIDA	OCOYOACAC
PAN	(9.1 , 10.0)	(49.2 , 50.6)	(8.5 , 10.06)
PRI	(61.1 , 62.5)	(47.6 , 48.0)	(35.8 , 39.1)
PRD	(25.0 , 26.4)	(0.3 , 0.5)	(26.3 , 29.3)
OTROS	(2.7 , 3.3)	(1.2 , 1.6)	(23.1 , 26.7)

Para analizar el comportamiento de las estimaciones a lo largo de la transmisión y captura de información, se fue acumulando y procesando la información cada vez que llega una nueva sección.²

² En el caso de Morelos y Mérida, dado que el sistema de captura con el que trabaja Opinión Profesional registró la hora a la que se capturó la información, se puede reproducir el desarrollo de la recepción de información. Sin embargo, para el levantamiento en el Estado de México no se contaba con ese dato, así que se ordenó la base de acuerdo al número de la sección y se fue añadiendo al proceso una sección a la vez para obtener un ejercicio similar.

Así, la primera estimación se generó con la información de la primera sección que se capturó, y la estimación final incluye todas las secciones que se transmitieron al centro de acopio.

En el proceso se tomaron en cuenta los factores de corrección al ponderador necesarios. Las gráficas resultantes de este ejercicio para cada uno de los operativos antes mencionados se presentan en las gráficas 4.1, 4.2 y 4.3.³

Si se observan estas gráficas pueden realizarse las siguientes afirmaciones:

A medida que la proporción de la característica de interés es menos prevalente en la población, se necesita una muestra acumulada mayor para que se establezcan las estimaciones de las proporciones.

Como se mencionó anteriormente, la precisión para un caso de competencia entre dos partidos o un pluripartidismo se establece de manera que los intervalos no se traslapen. Sin embargo, se observa que aun cuando ya se ha cumplido la totalidad de la muestra, los resultados que se generan en el caso de pluripartidismo no se cumple con esta condición.⁴

Esto se debe a que encuestas previas al levantamiento sugerían un comportamiento menos competido entre el PRI y el PRD. Y, dado que las proporciones de votos obtenidos el día de las elecciones están más cercanas entre sí que las de encuestas anteriores en el municipio, el error en las estimaciones es mayor que el que se quería. Por tal motivo, debe cuidarse muy bien este aspecto en la planeación de un AIRE y de toda encuesta en general. Sin embargo, debe también considerarse que cuando las proporciones están muy cercanas, la muestra se

³ Las tablas de resultados que apoyan estas gráficas se presentan en el Anexo B.

⁴ El hecho de que para realizar el ejercicio de ir acumulando las secciones se haya organizado la base de datos de acuerdo al número de sección, no afecta el último punto de la gráfica en donde se traslapan los intervalos. Porque, si bien quizá este no fue el comportamiento de la variación en la distribución de los votos, al final se cuenta con todas las secciones reportadas e independientemente del orden en que suceden éste es el resultado final generado.

incrementa mucho si se quiere una precisión alta, tanto, que en ocasiones sugiere casi un censo en la población. Como eso no es posible por el incremento en el costo, se deben balancear ambos aspectos en el diseño de la encuesta.

Por último, se observa que la longitud de los intervalos de confianza se hace más pequeña con el aumento en el tamaño de la muestra, porque al aumentar la muestra aumenta la precisión de las estimaciones, es decir, disminuye el error de muestreo y por lo tanto los intervalos de confianza son más angostos. Aunque, como se dijo antes, en los intervalos de confianza no se reflejan los niveles de los errores no de muestreo presentes en las estimaciones; sin embargo, dado el cuidado que se pone para tratar de minimizarlos, se consideran no significativos.

Para la presencia rara de la característica de interés se realizó un proceso análogo y en la gráfica 4.4 se presentan los resultados.

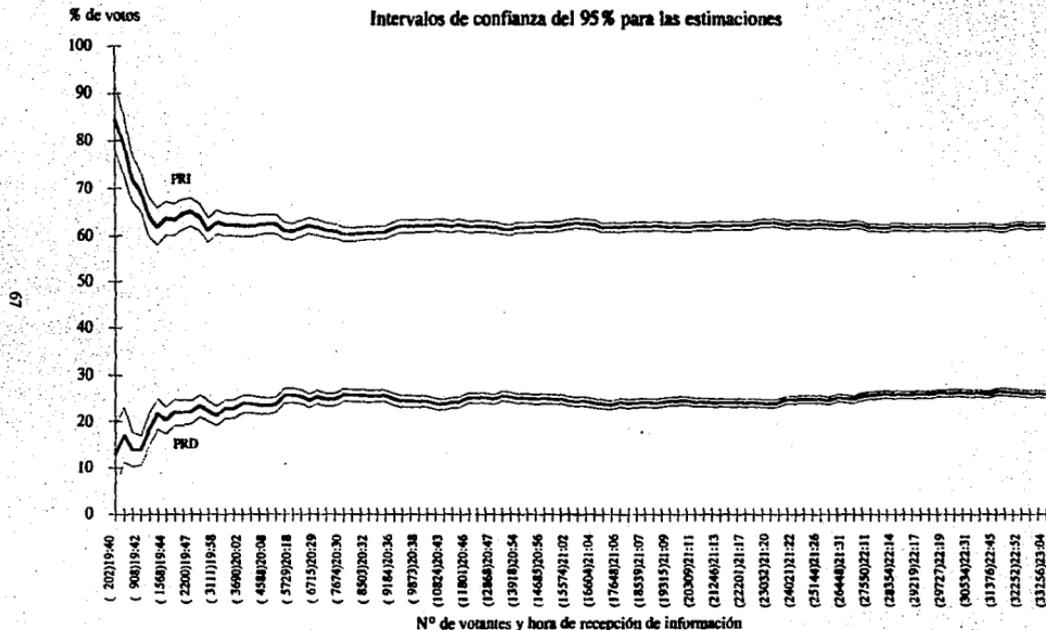
Además del decrecimiento de la longitud de los intervalos de confianza con el aumento de la muestra, en la gráfica 4.4, debido a la escala de la gráfica, nótese que prácticamente la estimación no sufrió cambios.

El tamaño de muestra de este levantamiento se calculó para generar resultados con precisión relativa de características con mayor presencia en la población, así que por tal motivo, con la muestra sólo se puede manifestar que el número o la proporción de votos para el PRD es una característica poco presente en la población. Esto es perfectamente válido, porque si se calculara un tamaño de muestra para obtener la proporción de votos para el PRD con una precisión relativa alta, el tamaño de muestra se incrementaría notablemente, sin repercutir en el mismo grado en la estimación de la primera y segunda fuerza electoral.

Gráfica 4.1

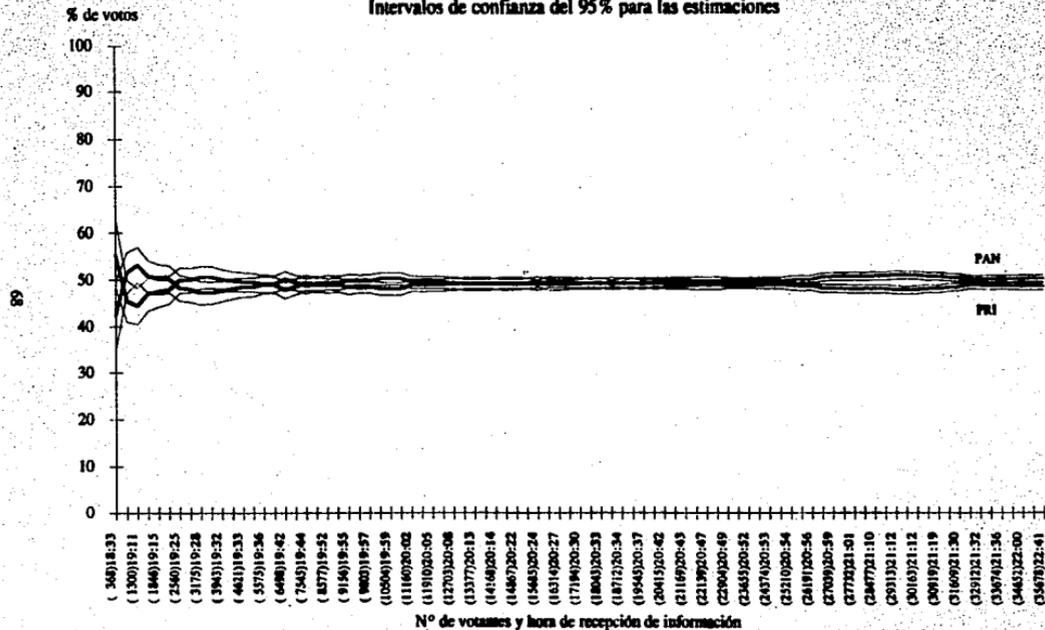
Elecciones para Gobernador en Morelos

Intervalos de confianza del 95% para las estimaciones



Gráfica 4.2

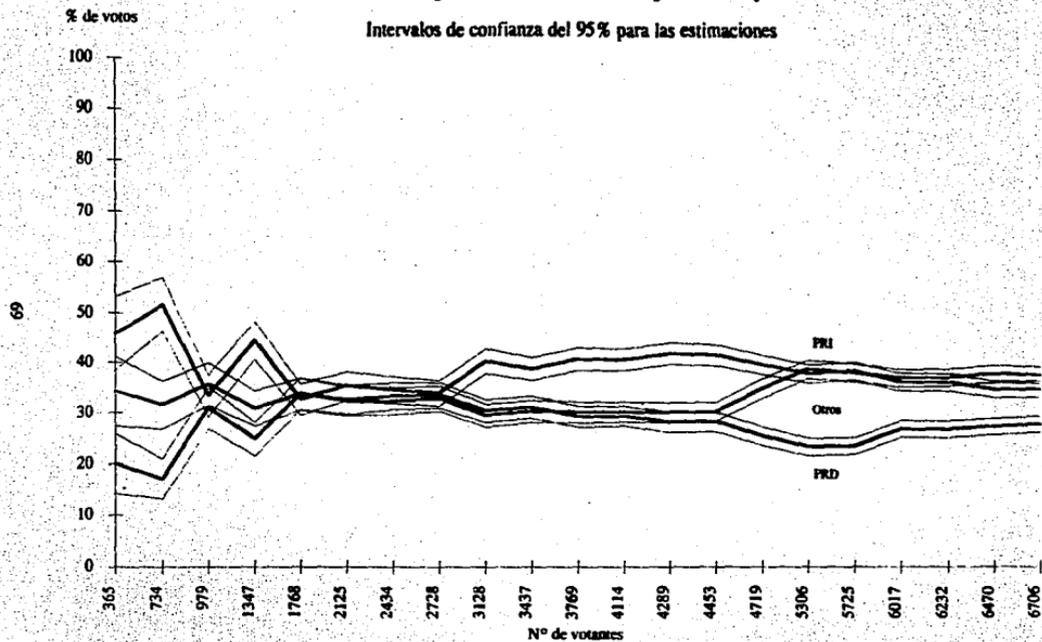
Elecciones para Presidente Municipal en Mérida
Intervalos de confianza del 95% para las estimaciones



Gráfica 4.3

Elecciones para Presidente Municipal en Ocoyoacac

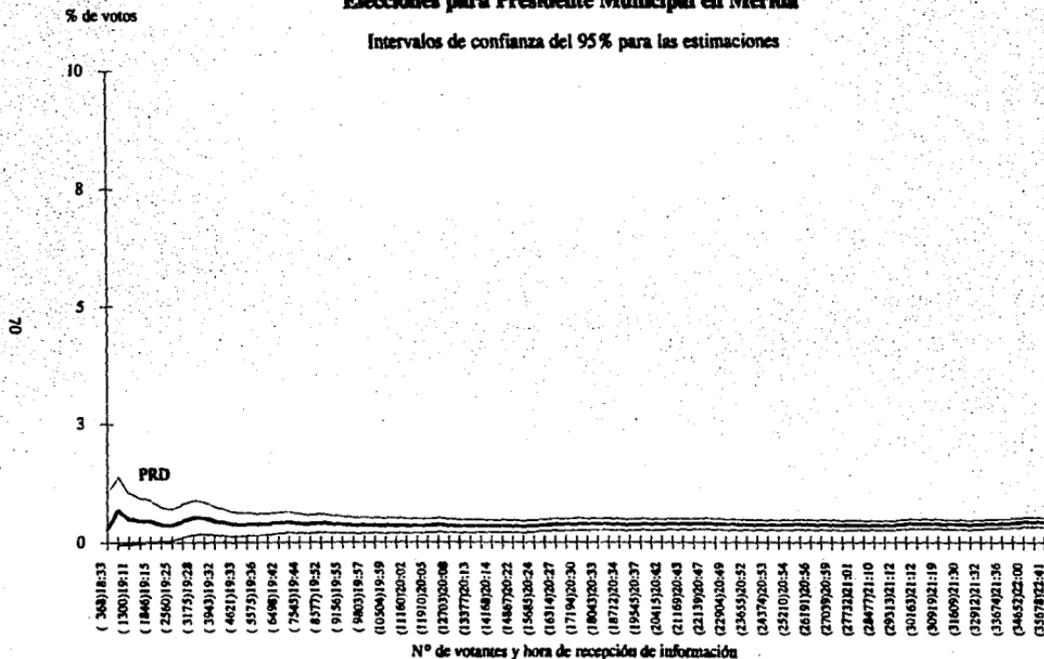
Intervalos de confianza del 95% para las estimaciones



Gráfica 4.4

Elecciones para Presidente Municipal en Mérida

Intervalos de confianza del 95 % para las estimaciones



Una vez visto el comportamiento de las estimaciones de proporciones en los diferentes niveles de presencia de la característica de interés, ahora, con estas encuestas se tratará de ejemplificar algunas razones del diseño de muestreo: la importancia de los estratos y el abarcar zonas urbanas y rurales.

Aunque el estudio de dos o tres levantamientos en particular no es suficiente para el sustento de una metodología, sí sirve para ejemplificar algunos aspectos de dicho soporte.

Para ilustrar las razones del uso de estratos, en cada uno de los levantamientos se procesó la información con un estrato a la vez. Los resultados obtenidos se presentan en el cuadro 4.5 además de la explicación del comportamiento de la preferencia electoral de las secciones que se agrupan en cada estrato.

Luego, se realizó una prueba de diferencia de proporciones y la prueba ji-cuadrada para funciones de distribución para calcular si existe una diferencia estadísticamente significativa entre las estimaciones de proporciones y, en conjunto, la distribución de los diferentes estratos (comparaciones entre cada uno de ellos). Además, se aplican estas pruebas para verificar la diferencia entre las estimaciones finales generadas con la información y los resultados obtenidos con cada uno de los estratos.⁵

⁵ Para el desarrollo de los cálculos ver Anexo C.

Cuadro 4.5
Procesamiento de la información en relación a los estratos⁶

MORELOS

	Todos los estratos	Estrato 1	Estrato 2	Estrato 3
PAN	9.5	6.1	16.0	0.5
PRI	61.8	65.3	57.6	58.3
PRD	25.7	25.5	23.2	40.3
OTROS	3.0	3.1	3.2	0.9
N° votantes	33526	17331	13992	2203

ESTRATO	91	94 ⁷	N° secciones
1	PRI	PRI (PRD)	39
2	PRI	PRI - PRD	24
3	PRI (PRD)	PRI - PRD	6

MERIDA

	Todos los estratos	Estrato 1	Estrato 2	Estrato 3
PAN	49.9	51.1	47.6	51.9
PRI	48.3	47.1	50.6	46.2
PRD	0.4	0.4	0.4	0.5
OTROS	1.4	1.4	1.4	1.4
N° votantes	35678	3019	8383	24241

ESTRATO	91	N° secciones
1	PRI	7
2	PAN (PRI)	13
3	PRI - PAN	41

OCOYOACAC

	Todos los estratos	Estrato 1	Estrato 2
PAN	9.5	8.1	12.0
PRI	37.5	41.4	30.2
PRD	27.8	25.4	32.3
OTROS	25.2	25.2	25.5
N° votantes	6706	4277	2429

ESTRATO	91	N° secciones
1	PRI	7
2	PRI mayoría relativa	4

⁶ El partido que se enuncia primero es el que predomina
 Cuando un partido se pone entre paréntesis significa que tiene una presencia relativa importante
 Cuando hay competencia entre dos partidos se separan por un guión
 Si se presenta un guión y paréntesis, la presencia de la segunda fuerza es importante

⁷ Información obtenida a través de encuestas pre-electorales.

En el AIRE en Morelos la diferencia de proporciones entre los estratos revela que no hay evidencia para decir que las secciones que componen cada uno de los estratos se comportan igual; en otras palabras, que cada uno de los estratos generan proporciones diferentes. Con la prueba ji-cuadrada se rechaza la hipótesis de que las distribuciones de los diferentes estratos sean iguales. En relación a la descripción de los estratos, nótese que las secciones que se quedaron en el estrato 1 conservaron su comportamiento electoral; las del estrato 2, en conjunto, generaron resultados un tanto diferentes a los que se utilizaron para su estratificación, aumentó la presencia del PAN y disminuyó la presencia del PRD, mientras que el PRI mantuvo su porcentaje de votos. Por su parte, el estrato 3 se concebía como un estrato con una presencia perredista más arraigada, misma que se confirmó en estas elecciones porque estas secciones generan un escenario competido entre el PRI y el PRD.

En el caso de Mérida se observa que cada uno de los estratos agrupa secciones que en su conjunto generan proporciones y distribuciones diferentes. Además, el comportamiento de 1991 se mantuvo en las elecciones de 1993, ya que cada uno de los estratos genera resultados según la descripción de su estrato.

Respecto al AIRE realizado en Ocoyoacac, como en los dos casos anteriores también se afirma que no hay evidencia para decir que las proporciones para cada uno de los partidos estimada por un estrato son iguales a las del otro estrato, y que entonces la distribución de los estratos es diferente. De acuerdo a la descripción de los estratos, el comportamiento cambia de prímico a pluripartidismo, con algunas diferencias entre ellos.

Con esto se ve que el hecho de formar estratos, en donde cada estrato representa un comportamiento diferente de la preferencia electoral, asegura el tener información de los diferentes comportamientos electorales. Aunque esta variación también se puede captar a través de un muestreo aleatorio simple, se opta por no dejar esta responsabilidad a este tipo de diseño de muestreo, ya que existe la probabilidad de omitir determinada preferencia

electoral o, en un caso extremo, considerar en la muestra sólo un tipo de sección, lo que sería un factor que induciría a errar en la estimación de la distribución del voto.

De una manera empírica, y a simple vista, en el cuadro anterior se observa que si la muestra hubiera considerado sólo un estrato en particular, para cualquiera de los escenarios, las estimaciones hubieran sido diferentes a las que se generaron con la información incluyendo las secciones de cada uno de los estratos. Con la prueba de diferencia de proporciones y la de ji-cuadrada se reafirma esta observación.

En un levantamiento de AIRE se involucran secciones en áreas urbanas y rurales. Para saber si la distribución de votos en las dos áreas difiere o no de manera significativa, se procesaron por separado las casillas urbanas y rurales para los operativos de Morelos y Ocoyoacac.⁸

Las estimaciones de la distribución de los votos en cada una de este tipo de casillas:

	MORELOS		OCOYOACAC	
	URBANA	RURAL	URBANA	RURAL
PAN	13.5	4.3	8.9	10.1
PRI	62.0	61.5	30.6	45.5
PRD	20.8	32.1	31.4	23.6
OTROS	3.6	2.1	29.1	20.8
Nº votantes	19,939	13,587	3,704	3,002

⁸ No se pudo realizar este ejercicio con la información obtenida en las elecciones en Mérida porque este municipio no cuenta con un número significativo de secciones rurales. En cuanto a Morelos, hay secciones que se clasifican como mixtas, pero de acuerdo a un estudio realizado por Opinión Profesional, este tipo de secciones se comporta de manera similar a las secciones rurales, así que para este ejercicio las secciones mixtas se tomaron como rurales.

A simple vista, se ve que las distribuciones de los votos de zonas urbanas son diferentes a las rurales. Sin embargo, puede hacerse una prueba de diferencia de funciones de distribución.⁹

La hipótesis es que la distribución de voto no difiere por tipo de área: urbana o rural. Una vez realizada la prueba de ji-cuadrada se observa que no hay evidencia para decir que las distribuciones de los votos son las mismas en zonas urbanas y zonas rurales. Por lo tanto, aunque pareciera conveniente levantar únicamente en zonas urbanas porque hay un más fácil acceso a medios de comunicación en éstas, así como costos menores, deben considerarse en la muestra secciones en las dos zonas.

Finalmente, se presenta la comparación entre los resultados del AIRE y los resultados oficiales.

En general, se observa que las estimaciones obtenidas del AIRE no distan mucho de los resultados oficiales. Sin embargo, en el caso de Ocoyoacac, se presenta una diferencia de 8.7 puntos porcentuales en la estimación de la proporción de votos obtenidos por el PRI. Anteriormente se mencionó que como las estimaciones resultantes del AIRE proponían un escenario más competido que el que mostraban algunas encuestas previas al levantamiento, se habían obtenido traslapes en los intervalos de confianza de las estimaciones de las proporciones de votos. Ahora, al comparar los resultados del AIRE y los oficiales, se observa que efectivamente las proporciones de votos obtenidos por los diferentes partidos no estaban tan cercanas, y se ven claramente las limitaciones del muestreo. A pesar de cuidar que el tamaño de la muestra y del diseño de la misma sean adecuados, al trabajar con una muestra de la población se corre el riesgo de obtener una muestra atípica.

⁹ Para el desarrollo de los cálculos ver Anexo C.

	AIRE	Oficiales	AIRE-Oficiales
MORELOS			
PAN	9.5	9	-0.5
PRI-PARM-PFCRN	61.8	67	5.2
PRD	25.7	22	-3.7
PPS-PT	3.0	2	-1.0

	AIRE	Oficiales	AIRE-Oficiales
MERIDA			
PAN	49.9	49.5	-0.4
PRI	48.3	50.5	2.2
OTRO	1.8	0.0	-1.8

	AIRE	Oficiales	AIRE-Oficiales
OCOYOACAC			
PRI	37.5	46.2	8.7
PRD	27.8	24.3	-3.5
OTRO	34.8	29.5	-5.3

4.2. EL ALCANCE DE UNA ENCUESTA POSVOTO

Tanto el AIRE como la EP tienen como objetivo estimar la distribución de los votos el día de las elecciones, y dado que ambos levantamientos tienen el mismo diseño de muestreo, los resultados del análisis del comportamiento de las estimaciones de proporciones con un AIRE son aplicables a una EP.

Por tal motivo, esta sección se enfoca al segundo objetivo de las EP: describir algunas características de los votantes, a partir de la EP realizada en Yucatán el 28 de noviembre de 1993 para obtener resultados al nivel estatal.¹⁰

Para recabar la información necesaria para cubrir este segundo objetivo de las EP se utiliza un cuestionario.

La preparación del cuestionario en una encuesta es una operación compleja y delicada. La naturaleza de las preguntas, la forma en la que están redactadas y el orden en que se suceden, tienen gran importancia para los resultados de la encuesta.

En un cuestionario se puede pedir a la persona interrogada que proporcione datos sobre un hecho o sobre una acción, o incluso, que indique su opinión en relación a temas específicos. De esta forma por su naturaleza, las preguntas se dividen en: preguntas de hecho o acción y preguntas de intención u opinión.

Las preguntas de hecho o de acción son numerosas en los censos: número de habitantes en la vivienda, lugar de origen, estado civil, etc. En las encuestas de opinión son necesarias cierto número de preguntas de hecho para la clasificación de las respuestas: sexo, edad, ocupación, etc.

Las respuestas a las cuestiones de hecho son fáciles y concretas. Estas pueden compararse con las preguntas de acción, en las que se pregunta si ha realizado determinado acto. Por ejemplo: "¿Por qué partido o candidato votó para gobernador?". En este caso la respuesta también puede ser concreta, porque la pregunta se refiere a algo concreto; la imprecisión puede provenir del alejamiento en el tiempo del acto respecto del momento de la encuesta.

¹⁰ Esta sección se basa en el uso de estadística descriptiva, ya que como en su momento se señaló, esta rama de la estadística involucra la representación gráfica de ciertos valores característicos, entre ellos las frecuencias, para elaborar términos descriptivos de lectura más simple y con una interpretación general.

Sin embargo, de no incluir los elementos metodológicos necesarios en ambas preguntas (de hecho o de acción), la desconfianza del entrevistado hace difícil la obtención de respuesta a algunas preguntas.

En cambio en las preguntas de intención, como su nombre lo indica, se pregunta cómo actuaría si eventualmente se presentara una situación específica. La respuesta es más difícil, porque es más fácil acordarse de una acción real que imaginar una acción teórica.

Además, la respuesta de una pregunta de intención debe interpretarse como reveladora de la opinión del individuo entrevistado, pero no como manifestación segura de la acción que realizaría si efectivamente debiera actuar. Por ejemplo, la pregunta: "Imagine que el día de hoy son las elecciones, ¿por qué partido votaría usted?", es útil para conocer las opiniones políticas del individuo entrevistado en el momento de la encuesta, pero, si efectivamente hubiera elecciones, la campaña electoral y otras circunstancias quizá le harían votar en sentido diferente al indicado.

Así pues, las preguntas de intención más bien se asemejan a las preguntas de opinión. "¿Qué es lo primero que debe realizar el próximo gobierno?", es una cuestión de opinión. La pregunta de opinión pide al individuo que diga lo que piensa, no lo que hace o lo que sabe. La respuesta es más difícil y menos precisa que en las preguntas de hechos o acontecimientos, porque es más difícil delimitar y expresar las opiniones que los hechos o los acontecimientos.

Aunque cabe señalar que, no importa la naturaleza de las preguntas, si éstas se refieren a aspectos que el entrevistado considera más personales, generalmente se corre un mayor riesgo de no respuesta o de obtener respuestas falsas.

En cuanto a la libertad de respuesta, existen preguntas cerradas y abiertas. En las preguntas cerradas, ya se tienen predeterminadas las diferentes categorías de respuesta; en cambio en las preguntas abiertas se registra la respuesta tal y como la expresa el entrevistado. Ambas tienen ventajas y desventajas. Si bien las preguntas abiertas permiten percibir más claramente la actitud y las opiniones de los individuos, existen problemas en el momento de la interpretación y codificación de las respuestas, además de ser un proceso más lento que el de las preguntas cerradas. Aunque, si bien es cierto que las preguntas cerradas no causan mayor problema en la codificación e interpretación, es válido mencionar que, dadas las opciones de respuesta, al responder el entrevistado se vio "forzado" a adaptar su respuesta a alguna categoría predeterminada, aunque ésta no representara exactamente lo que él hubiera respondido.

Al realizar un análisis matemático de resultados de una encuesta, se parte de que los fenómenos en estudio son mensurables o aproximadamente mensurables.

Ejemplos de fenómenos mensurables son: datos demográficos, fenómenos económicos y un gran número de elementos geográficos. "En la vida política, las elecciones constituyen el tipo exacto de fenómeno materialmente mensurable."¹¹

Entre los fenómenos aproximadamente mensurables se encuentra la medición de actitudes o de las opiniones. La dificultad de apreciar la intensidad de una actitud o de una opinión reside en la interpretación porque

a pesar del esfuerzo realizado para establecer unas escalas objetivas y fieles, siempre queda un amplio coeficiente de subjetividad. Esto no quiere decir que las cifras estén desprovistas de toda significación, sino únicamente que constituyen órdenes de magnitud, simples indicaciones, y que, por consiguiente, no deben ser consideradas con rigor.¹²

¹¹ Duverger, Maurice. *Métodos de las Ciencias Sociales*. Ariel, México, 1981. p.440

¹² *Ibidem*, p. 441

Con base en las 17 preguntas del cuestionario de la EP realizada en Yucatán, como 13 de ellas se pueden clasificar como de hecho o de acción; se puede decir que la información que se captó en esta EP es concreta y sin muchos aspectos subjetivos.¹³ A partir de estas preguntas se pueden realizar cruces con la preferencia electoral y así, además de conocer los resultados globales, se pueden obtener en el nivel de partido.

Después, se comparan los resultados de las preguntas socioeconómicas con la información del último censo de población y vivienda con las personas de 18 años y más, a fin de comparar la participación de la población en las votaciones. En el cuadro 4.6 se presentan los resultados obtenidos.

Dado que la selección de los entrevistados se considera un proceso aleatorio, se espera que la información recabada en una EP se parezca a los parámetros calculados a partir del XI Censo Nacional de Población y Vivienda.

De la comparación de los porcentajes obtenidos en la EP de Yucatán y los parámetros calculados a partir del censo, se puede observar que:

1. Hubo una mayor participación de hombres, ya que siendo ellos el 48.8% de la población, el 56.2% de los votantes fueron de sexo masculino.
2. Los adultos (26 a 60 años) acudieron en mayor porcentaje a emitir su voto.
3. Aquellas personas con secundaria incompleta y más, participaron en mayor proporción en las elecciones para gobernador.
4. Entre los que no forman parte de la población económicamente activa, amas de casa, estudiantes, jubilados y desocupados, fueron los jubilados y desocupados quienes en mayor proporción acudieron a votar.

¹³ El cuestionario que se aplicó en la EP en Yucatán está en el Anexo D.

Cuadro 4.6
Comparación de resultados obtenidos en el XI Censo de Población y
la Encuesta Posvoto en Yucatán (28 de noviembre de 1993)

	Parámetros del XI Censo	Encuesta Posvoto
SEXO		
Femenino	51.3	43.8
Masculino	48.8	56.2
EDAD		
18 a 25 años	27.0	21.8
26 a 40 años	35.4	40.2
41 a 60 años	24.9	28.4
61 años y más	12.6	9.6
ESCOLARIDAD		
Ninguna	16.5	11.2
Primaria	50.2	43.4
Secundaria o preparatoria	27.7	30.4
Universidad y más	5.6	15.0
POB ECONOMICA INACTIVA		
Ama de casa	74.4	70.5
Estudiante	18.6	11.9
Jubilado	5.5	11.9
Desocupado	1.4	5.7

A partir de estos resultados, y también considerando las preguntas de opinión, a la fecha se ha intentado describir al votante mexicano con algunos estudios exploratorios pues no existen en México encuestas realizadas con este fin específico.

Finalmente, con la intención de enriquecer la información que se puede generar con las EP, a continuación se presentan los valores numéricos correspondientes al flujo de votantes (el porcentaje de votantes en relación a la hora de la emisión de su voto) global y por preferencia electoral de las dos primeras fuerzas contendientes, así como su representación gráfica, cuadro 4.7 y gráfica 4.8 respectivamente.

Cuadro 4.7
% de votantes en relación a la hora de entrevista

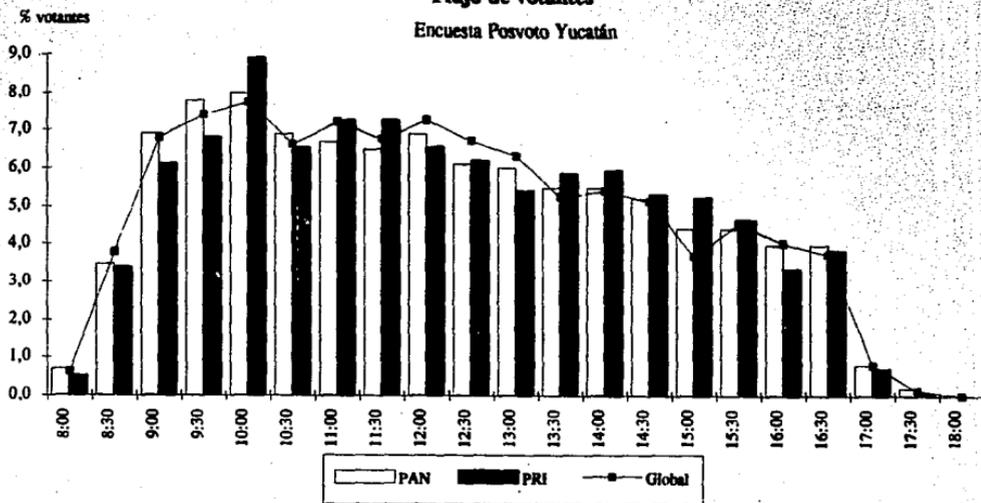
Hora de entrevista	% de votantes global	% acumulado de votantes	% de votantes PAN	% de votantes PRI
8:00	0,7	0,7	0,5	0,6
8:30	3,5	4,2	3,4	3,8
9:00	6,9	11,1	6,1	6,8
9:30	7,8	18,9	6,8	7,4
10:00	8,0	26,9	8,9	7,7
10:30	6,9	33,8	6,6	6,6
11:00	6,7	40,5	7,3	7,2
11:30	6,5	47,0	7,3	6,8
12:00	6,9	53,9	6,6	7,3
12:30	6,1	60,0	6,2	6,7
13:00	6,0	66,0	5,4	6,3
13:30	5,5	71,5	5,9	5,3
14:00	5,5	77,0	5,9	5,4
14:30	5,2	82,2	5,3	5,1
15:00	4,4	86,6	5,2	3,7
15:30	4,4	91,0	4,6	4,5
16:00	4,0	95,0	3,3	4,1
16:30	4,0	99,0	3,8	3,7
17:00	0,8	99,8	0,7	0,8
17:30	0,2	100,0	0,1	0,1
18:00	0,0	100,0	0,0	0,0

* Los renglones sombreados señalan la hora a la que se acumula aproximadamente el 25, 50, 75 y 100% de los votantes según la hora de entrevista registrada en los cuestionarios de la Encuesta Posvoto.

Gráfica 4.8

Flujo de votantes
Encuesta Posvoto Yucatán

83



En el cuadro 4.7 se observa que aproximadamente a la mitad de la jornada electoral, 1:30 a 2:00 p.m., habían votado dos terceras partes del total de personas que irían a emitir su voto en las elecciones para gobernador de Yucatán. Por otra parte, en la gráfica 4.8, el flujo de votantes por preferencia electoral refleja una asistencia a las urnas muy similar entre los partidarios del PAN y PRI.

A continuación se presentan los resultados generados por EP, AIRE y los resultados oficiales.

	EP	AIRE	Oficiales	EP-Oficiales	AIRE-Oficiales
YUCATAN					
Gobernador					
PAN	32.7	40.3	37.9	-5.2	2.4
PRI	63.7	58.7	60.1	3.6	-1.4
PRD	3.6	1.1	2.0	1.6	-0.9

MERIDA

Pte Municipal

PAN	42.2	49.9	49.5	-7.3	0.4
PRI	52.6	48.3	50.5	2.1	-2.2
PRD	5.2	1.8	0.0	5.2	1.8

Como se puede observar en el cuadro anterior, aunque los resultados de las EP se asemejan a los oficiales, existe una menor diferencia entre estos últimos y los obtenidos por un AIRE.

Para este levantamiento en particular, se realizó EP y AIRE. Se consideró conveniente enfocar el AIRE para la estimación de la distribución de los votos y la EP para describir el perfil del votante. Esto permitió disminuir el tamaño de muestra de la EP, ya que se calculó

para un precisión menor, debido a que el costo por sección es mayor para una EP que para un AIRE.

Luego, para ampliar la argumentación de las diferencias entre las estimaciones de los dos levantamientos se presenta la siguiente información.

EP	YUCATAN	MERIDA
	%	%
Votos anulados	2.5	0.9
Votos para ninguno de los partidos	0.5	0.2
Respuesta no válida	3.0	1.1
Votos en blanco	14.9	22.8
Entrevistados que no pudieron votar	1.8	2.0
Cuestionarios incompletos	2.4	2.1
No respuesta en preferencia electoral	19.1	26.9
AIRE		
Respuesta no válida	0.7	0.4
No respuesta en preferencia electoral	1.4	1.4

Estas cifras son importantes pues de rebasar la magnitud planeada, no se podrían obtener estimaciones confiables a partir de una EP, porque estos valores constituyen la no respuesta y entonces la tasa de respuesta, que es el cociente del número de entrevistas y el tamaño de la muestra, se reduce de manera considerable.

La no respuesta puede tener una influencia relevante en las estimaciones. Como se observa en el cuadro inmediato anterior, en comparación con el AIRE, en la EP existe un mayor porcentaje de casos de los que no se puede obtener la preferencia electoral.

En general, existen tres formas de manejar el problema de la no respuesta:¹⁴

1. Seleccionar una muestra aleatoria entre los entrevistados que no respondieron, tratar de contactarlos y volver a aplicarles el cuestionario. Se dice que este es el método más aconsejable. Sin embargo, como los votantes se entrevistan afuera de las casillas el día de la elección y se les aplica un cuestionario *anónimo*, no es factible contactarlos a futuro.
2. Estimar los posibles valores extremos (o promedios) de la característica de interés en la población que no responde y construir escenarios usando estos valores. Esta aproximación introduce un problema serio: las estimaciones dependen de los supuestos usados para estimar los valores extremos o los promedios, y estos valores son producto de conjeturas. En particular en México, a este momento, no hay estudios del votante que permitan definir un criterio general, aplicable a situaciones establecidas, para la asignación de no respuesta en cuanto a encuestas para medición de preferencias electorales.
3. Simplemente presentar los resultados junto con la información de no respuesta y dejar que el lector interprete los resultados. Esta es una manera sencilla de tratar la no respuesta, pero si lo que interesa es saber quién ganó las elecciones la noche misma de la votación, la información relativa a la no respuesta no es útil en ese momento.

En realidad, el problema de la no respuesta se centra en saber si las personas que no respondieron tienen un comportamiento específico con respecto a la característica de interés o si las condiciones que inducen a la no respuesta no se involucran directamente con dicha

¹⁴ Futrell, David. "Ten Reasons why surveys fail". *Quality Progress*, Vol. 27, Núm. 4, Abril 1994, p. 66

característica. En otras palabras, en el caso práctico de la medición de preferencias electorales, la no respuesta puede estar relacionada con el partido por que el votaron o puede no estarlo, si lo está y la no respuesta es alta se estaría subestimando a al partido que se relaciona con la no respuesta, en cambio si no lo está, un porcentaje alto de no respuesta afectaría en la precisión de las estimaciones en general, pero no modificaría la tendencia de las mismas. Nótese entonces que la no respuesta en sí, es un problema complejo.

Dado que la no respuesta, en términos prácticos "reduce" la muestra, esto implica que también se reduce la precisión de los resultados. En el AIRE hay menor porcentaje de no respuesta y además, dado que la información recolectada para la casilla corresponde a la totalidad de votantes, no hay error muestral en esta última etapa. Por lo tanto, el AIRE genera estimaciones más cercanas a los resultados oficiales, con los respectivos límites consecuentes de estimar los resultados a partir de una muestra.

En promedio, Opinión Profesional tiene una diferencia de 1.8% en sus estimaciones a partir de AIRE con los resultados oficiales comparada con un 2.6% con la información recabada en una EP.

CONCLUSIONES

En la planeación de una encuesta, la teoría del muestreo con la que se cuenta permite el cálculo del tamaño de muestra considerando la precisión de las estimaciones e involucrando el nivel de confianza de las mismas.

Además, esta teoría considera la realización de diseños de muestreo complejos, los que pueden constituir encuestas con operativos de campo más prácticos, costos menores, adecuación al marco de muestreo disponible y estimaciones más precisas; todos estos aspectos importantes para el cumplimiento de los objetivos de una encuesta por muestreo.

En relación a los resultados generados de una muestra, si la muestra es probabilística, se pueden hacer inferencias acerca de la población de donde se obtuvo la muestra.

En particular, el muestreo de proporciones, además de gozar de los beneficios del muestreo como técnica especializada, permite el uso de una medida de localización muy difundida: la proporción o en su defecto, el uso de porcentajes.

Sin embargo, a pesar de los alcances del muestreo, como toda técnica tiene limitaciones y una de ellas se manifiesta en los errores no de muestreo, éstos no pueden cuantificarse ni controlarse con el aumento del tamaño de muestra como se hace con los errores de muestreo. Luego, las estimaciones que se generan y que presumen cierta precisión no están considerando los errores no de muestreo en los que se incurrió durante el levantamiento de la encuesta. Pero, cabe mencionar que los expertos en esta área al darse cuenta de esto, han introducido en el desarrollo de una encuesta medidas para minimizar este tipo de errores.

Entre estas medidas se pueden enunciar: diseño de un operativo de campo práctico, definición de criterios de codificación generales, limpieza de archivos capturados, supervisión en los diferentes pasos del manejo de la información recabada en la encuesta y, por supuesto, capacitación a los encuestadores para, además de promover una correcta aplicación del cuestionario y evitar su influencia en las respuestas del entrevistado, tratar de reducir la no respuesta. En esta labor el diseño del cuestionario es un punto esencial.

Otro aspecto importante es no olvidar que las estimaciones se realizan a partir de los resultados de una muestra, y que cuando se especifica un nivel de confianza de 95%, se espera que de 100 muestras que se saquen, en 95 de ellas los intervalos de confianza contendrán el valor poblacional; cuando se realiza una encuesta sólo se trabaja con una muestra y existe la probabilidad de que ésta sea atípica. Tal vez se opine que lo que se debe hacer entonces es aumentar el nivel de confianza a un 99% pero, en este caso, la longitud de los intervalos también aumenta y como en los demás aspectos del muestreo que están en conflicto, el experto en muestreo debe equilibrarlos para obtener un "buen" diseño de muestreo.

Este trabajo permite ver el cumplimiento de algunos aspectos de la teoría de muestreo:

En el cálculo del tamaño de muestra, la importancia de conocer la dispersión del fenómeno de estudio, ya que cuanto más dispersos estén los valores de la variable en estudio, la muestra que se requiere es mayor.

La necesidad de considerar los diferentes tipos de diseños de muestreo para que, de acuerdo a los objetivos de la encuesta, las estimaciones que se quieran realizar, el costo y el marco de muestreo disponible, den lugar a una buena muestra.

El decrecimiento de los intervalos de confianza a medida que aumenta el tamaño de muestra, porque al aumentar el tamaño de muestra, aumenta la precisión y por lo tanto, disminuye el error de muestreo.

Finalmente, se observó la estabilización de las estimaciones en relación al nivel de presencia de la característica de interés, precisión y tamaño de muestra.

En relación a la aplicación concreta de la teoría de muestreo para proporciones en la medición de preferencias electorales se observa que el procedimiento de selección que utiliza la empresa Opinión Profesional se adecua a la disponibilidad de información; se basa en estudios que han realizado para ver las diferencias en la preferencia electoral en áreas urbanas y rurales, distribución de la afluencia de votantes en las diferentes casillas que conforman una sección, etc.

Si bien es cierto que esta múltiple consideración de aspectos se debe a la experiencia que esta empresa tiene en este ramo, la aplicación de estos levantamientos y el surgimiento de nuevas situaciones no consideradas con anterioridad han hecho que estos operativos tengan una metodología compleja y comparable a la que se utiliza en otras partes del mundo. De esta manera, su continua aplicación y la disponibilidad de mayor información provocarán mejoras en el desarrollo de estos levantamientos.

La información que generan tanto el AIRE como la EP es de gran interés, porque se puede considerar como un instrumento de verificación de la limpieza electoral. Además, se puede describir la participación de los diferentes grupos en las elecciones y realizar un estudio comparativo de ésta con los resultados obtenidos en varios comicios con escenarios electorales semejantes; esto podría enriquecer la información que hasta ahora se tiene del votante mexicano.

Dada la estructura del presente trabajo, no es posible obtener resultados generales en cuanto a la descripción de las elecciones en México. En cambio, permite hacer algunos comentarios sobre los casos particulares que se analizaron.

En las elecciones de marzo de 1994 en Morelos, en donde predominó la presencia del PRI, las estimaciones se estabilizaron alrededor de las 21 horas, lo que quiere decir que más o menos a esta hora ya se hubiera podido decir con una precisión cercana a la planeada, quién y con qué porcentaje ganó las elecciones. Con respecto a los cambios en las preferencias electorales, en estas elecciones hubo una mayor presencia de partidos de oposición en comparación a 1991; el PAN aumentó ligeramente su presencia en relación a una encuesta pre-electoral que levantó Opinión Profesional en este Estado. Además, a partir de las estimaciones por tipo de zona, se puede decir que en Morelos, en dichas elecciones, el PRD tuvo más seguidores en las zonas rurales y el PAN en las áreas urbanas, mientras que los simpatizantes del PRI no presentaron diferencias por el tipo de zona.

En Mérida, las elecciones de noviembre de 1993 estuvieron muy cerradas, por lo que las estimaciones se estabilizaron hasta las 22 horas, cuando ya sólo faltaban 2 casillas de un total de 89 en muestra. En este municipio cada uno de los dos partidos "fuertes" conservó a sus simpatizantes, y como consecuencia no hubo cambios con respecto a la información que se recabó en 1991.

Es importante resaltar que la hora en la que se terminó de recibir información para cada uno de los levantamientos anteriormente citados no es comparable directamente, ya que algunos factores que intervienen en la duración de la recepción de información, además del tamaño de la muestra, son: el nivel de presencia de la característica de interés, el tipo de zona en la que se localizan las secciones en muestra, los medios de transmisión y las condiciones de seguridad en las que trabaja el recolector.

En Ocoyoacac, el AIRE que se realizó en los comicios que tuvieron lugar en noviembre de 1993 presentó un escenario más competido de lo que en realidad fue. Además, en este operativo no se contó con la información de hora de recepción, por lo que no se puede hablar ciertamente del comportamiento de las estimaciones en el transcurso de la recepción de la información. Sin embargo, de acuerdo a las estimaciones del AIRE, en Ocoyoacac hubo más priístas en las zonas rurales, más perredistas en las urbanas, mientras que los panistas se encontraron en igual porcentaje en ambas áreas.

En relación a la EP realizada en Yucatán en noviembre de 1993, en donde hubo predominio del PRI, con base en la información que es comparable con los parámetros obtenidos a partir del último censo, se observa que hubo una mayor participación en las votaciones por parte de los hombres, de ciudadanos que pertenecen al grupo de edad de 26 a 60 años, de personas con escolaridad al menos de secundaria incompleta, y entre la población inactiva hubo mayor participación por parte de los jubilados y los desocupados. En Yucatán, la población prefirió ir a votar temprano; a la mitad de la jornada electoral ya habían votado dos terceras partes de las personas que asistirían a votar. Sin embargo, no se percibe diferencia en el flujo de votantes entre panistas y priístas en el transcurso de la jornada electoral.

BIBLIOGRAFIA

- AZORIN Poch, Francisco. *Curso de Muestreo y Aplicaciones*. Aguilar. España, 1972. 325 p.
- BLALOCK, Hubert M. *Social Statistics*. 2 ed. McGraw-Hill. New York, 1979. 625 p.
- COCHRAN, William G. *Técnicas de Muestreo*. Tr. de Andrés Sestier Bouquier. C. E. C. S. A. México, 1980. 513 p.
- CORNFIELD, Jerome. "Modern Methods in Sampling of Human Populations. The Determination of Sample Size". *American Journal Publicacion Health*, 41. p.654-661.
- DUVERGER, Maurice. *Métodos de las Ciencias Sociales*. 12 ed. Tr. de Alfonso Sureda. Ariel. México, 1981. 593 p.
- FOWLER, Floyd J. Jr. *Survey Research Methods*. 2 ed. SAGE Publications. Newbury Park, 1993. (Applied Social Research Methods Series). 156 p.
- FREUND, John E. *Statistics. A First Course*. Prentice-Hall. New Jersey, 1970. 340 p.
- FUTRELL, David. "Ten reasons why surveys fail". *Quality Progress*. Vol. 27, Núm. 4, Abril 1994. p. 66
- GROVES, Robert M. *Survey Errors and Survey Costs*. John Wiley & Sons. New York, 1979. 590 p.
- HANSEN, Morris H. *et al. Sample Survey Methods and Theory Vol. 1*. John Wiley & Sons. New York, 1953. 638 p.
- KISH, Leslie. *Muestreo de Encuestas*. Tr. de Ricardo Vinós Cruz López. Trillas. México, 1972. 739 p.
- KREYSZIG, Erwin. *Introducción a la Estadística Matemática. Principios y Métodos*. Tr. de Arturo Galán Martínez. Limusa. México, 1987. 505 p.
- MENDENHALL, William, *et. al. Estadística Matemática con Aplicaciones*. Tr. de Dirk Valckx Verbeeck y Arturo de la Fuente Pantoja. Grupo Editorial Iberoamérica. México, 1986. 751 p.

- MERCADO Gasca, Lauro. *Estudio del Votante Mexicano por medio de las Encuestas a la Salida de las Casillas. Tesis.* El Colegio de México, Centro de Estudios Internacionales. México, 1993. 119 p.
- MILLER, Irwin R., et al. *Probabilidad y Estadística para Ingenieros.* 4 ed., Tr. de Virgilio González Pozo. Prentice-Hall Hispanoamericana. México, 1992. 624 p.
- MOOD, Alexander M., et al. *Introduction to the Theory of Statistics.* 3 ed. McGraw-Hill. Singapur, 1974. 564 p.
- OPINION PROFESIONAL. *Acopio Inmediato de Resultados Electorales por Muestreo. Nota Metodológica.* México. Julio, 1993.
- RODRIGUEZ Vera, Angel. *Determinación de Tamaño de Muestra mediante Programación No-Lineal. Tesis.* Facultad de Ciencias. U. N. A. M. México, 1975. 53 p.
- SCHEAFFER, Richard L. et al. *Elementos de Muestreo.* Tr. de Gilberto Rendón Sánchez. México. Grupo Editorial Iberomérica. México, 1987. 321 p.
- Manual del Sistema de Proceso de Información de Resultados Electorales.* Trabajo no publicado. Opinión Profesional. 1994.

FORMULAS PARA LA ESTIMACION DE PROPORCIONES Y SUS VARIANZAS A PARTIR DE DIFERENTES DISEÑOS DE MUESTREO

• Muestreo Aleatorio Simple

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

donde

$y_i = 1$ si tiene la característica de interés

$y_i = 0$ en otro caso

$$v(p) = \frac{pq}{n-1}(1-f),$$

donde

$$q = 1 - p$$

$$f = n/N$$

• Muestreo Aleatorio Estratificado

$$p_n = \sum_{h=1}^L W_h p_h,$$

donde

$$p_h = \frac{\sum_{i=1}^n y_{hi}}{n}$$

proporción de unidades en la muestra que pertenecen a una clase definida en el h -ésimo estrato

- $W_h = \frac{N_h}{N}$ peso relativo del h -ésimo estrato
 N_h número total de unidades en el h -ésimo estrato
 n_h número de unidades en muestra en el h -ésimo estrato
 y_{hi} valor obtenido para la i -ésima unidad en el h -ésimo estrato
 L número de estratos

$$v(p_u) = \sum_{h=1}^L W_h^2 \frac{p_h q_h}{n_h - 1} (1 - f_h)$$

• **Muestreo Sistemático**

El muestreo sistemático constituye una alternativa de la selección aleatoria e independiente de unidades de muestreo, y a veces se le llama pseudoaleatorio. En sentido estricto para que las varianzas de las muestras sistemáticas sean válidas los cálculos requieren que se realicen dos selecciones por zona.

$$v(p_{ms}) = \frac{1-f}{2n(n-1)} \sum_{s=1}^{n-1} (p_s - p_{s+1})^2$$

En la fórmula de la varianza se emplean $(n-1)$ diferencias sucesivas $(p_s - p_{s+1})$, donde el subíndice denota el orden de selección.

Pero como las muestras sistemáticas generalmente constan de selecciones únicas de cada zona se suelen ocupar los mismos estimadores de una muestra aleatoria simple, los que en condiciones normales dan buenas estimaciones.

• **Muestreo por Conglomerados en una Etapa**

$$p = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i},$$

donde

- a_i número total de elementos en el i -ésimo conglomerado que poseen la característica de interés
- m_i número de elementos en el i -ésimo conglomerado
- n número de conglomerados seleccionados en muestra

$$v(p) = \frac{N-n}{Nn(M/N)^2} \frac{\sum_{i=1}^n (a_i - pm_i)^2}{n-1},$$

donde

N número de conglomerados en la población

$M = \sum_{i=1}^N m_i$ número de elementos en la población

• **Muestreo por Conglomerados Desiguales en dos Etapas**

$$p = \frac{N}{M} \frac{\sum_{i=1}^N M_i p_i}{n},$$

donde

M_i tamaño del i -ésimo conglomerados

$p_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$ proporción muestral para el i -ésimo conglomerado

Si se desconoce M se puede sustituir por $N \sum M_i / n$ donde $\sum_{i=1}^n M_i / n$ es el tamaño de conglomerado promedio.

$$v(p) = \frac{N-n}{N} \frac{1}{n(M/N)^2} \frac{\sum_{i=1}^n M_i^2 (p_i - p)^2}{n-1} + \frac{1}{nN(M/N)^2} \sum_{i=1}^n M_i^2 \frac{M_i - m_i}{M_i} \frac{p_i q_i}{m_i - 1}$$

Un caso particular es cuando los conglomerados son iguales.

• Muestreo con Probabilidad Proporcional al Tamaño

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

donde

z_i probabilidad de que la i -ésima unidad sea seleccionada

$$v(p) = \frac{1}{N^2 n(n-1)} \sum_{i=1}^n z_i \left(\frac{y_i}{z_i} - Np \right)^2$$

ANEXO B

INTERVALOS DE CONFIANZA PARA LAS ESTIMACIONES PONDERADAS
DE LA DISTRIBUCION DEL VOTO EN ELECCIONES PARA GOBERNADOR
MORELOS

Nº VOTOS	HORA	(PRI)	(PRD)	(OTROS)	TIPO	ESTRATO
202	19:40	77,6	84,7	91,7	6,3	12,9	19,4	-0,6	2,5	5,5	R	1
317	19:41	72,9	79,3	85,6	11,2	17,0	22,9	0,8	3,7	6,7	R	1
664	19:42	67,2	72,0	76,9	10,2	14,0	17,7	10,2	14,0	17,7	U	2
908	19:42	65,1	69,4	73,6	10,7	13,9	17,0	13,3	16,8	20,2	U	1
1062	19:43	59,7	63,8	67,9	15,1	18,4	21,7	14,5	17,8	21,1	R	2
1239	19:43	58,0	61,9	65,7	18,4	21,6	24,9	13,6	16,5	19,5	M	3
1568	19:44	60,2	63,6	67,0	17,4	20,2	23,1	13,6	16,2	18,8	U	3
1765	19:44	60,2	63,4	66,6	19,2	21,9	24,7	12,3	14,7	17,0	R	3
1878	19:45	61,3	64,4	67,5	19,2	21,9	24,5	11,5	13,7	15,9	R	4
2200	19:47	62,1	65,0	67,8	19,5	22,0	24,4	11,0	13,1	15,1	R	5
2420	19:55	61,1	63,8	66,5	21,0	23,4	25,8	10,9	12,8	14,7	U	2
2838	19:56	58,6	61,2	63,7	19,9	22,1	24,3	14,7	16,7	18,7	U	2
3111	19:58	60,4	62,9	65,3	19,2	21,3	23,3	14,1	15,9	17,8	U	3
3326	19:59	59,9	62,3	64,6	20,5	22,6	24,6	13,4	15,2	16,9	M	3
3539	20:02	60,1	62,4	64,6	20,7	22,6	24,6	13,3	15,0	16,7	R	1
3690	20:02	59,8	62,0	64,2	21,7	23,7	25,7	12,7	14,3	15,9	R	5
3957	20:03	59,9	62,1	64,2	21,8	23,7	25,6	12,7	14,2	15,8	U	3
4116	20:07	60,2	62,3	64,4	21,6	23,5	25,3	12,8	14,3	15,8	M	2
4588	20:08	60,5	62,5	64,5	21,5	23,2	24,9	12,9	14,3	15,8	M	2
4958	20:08	60,6	62,6	64,5	21,9	23,6	25,2	12,5	13,9	15,3	M	4
5390	20:11	59,2	61,1	62,9	23,8	25,4	27,1	12,2	13,5	14,8	U	3
5729	20:18	59,1	60,9	62,7	24,0	25,6	27,2	12,3	13,5	14,8	U	2
6139	20:18	59,7	61,4	63,2	23,8	25,3	26,8	12,1	13,3	14,5	U	3
6412	20:19	60,5	62,2	63,9	22,8	24,3	25,8	12,3	13,5	14,7	U	3
6715	20:29	60,1	61,7	63,4	23,8	25,3	26,7	11,9	13,0	14,1	M	2
6996	20:30	59,6	61,2	62,8	23,1	24,5	26,0	13,1	14,3	15,5	U	3
7327	20:30	59,4	61,0	62,6	23,3	24,7	26,1	13,2	14,3	15,4	M	2
7674	20:30	58,7	60,3	61,8	24,2	25,6	27,0	13,0	14,1	15,2	R	1
7954	20:31	58,8	60,3	61,8	24,2	25,5	26,9	13,0	14,1	15,2	U	3
8209	20:31	58,8	60,3	61,8	24,2	25,6	26,9	13,0	14,1	15,2	M	2
8503	20:32	59,1	60,6	62,1	24,0	25,3	26,6	13,1	14,1	15,2	R	1
8742	20:33	59,1	60,6	62,0	24,2	25,5	26,8	12,9	13,9	15,0	M	2
8932	20:33	59,3	60,7	62,1	24,1	25,4	26,6	12,9	14,0	15,0	M	3
9184	20:36	60,2	61,6	63,0	23,4	24,7	25,9	12,8	13,8	14,8	R	3
9404	20:37	60,6	62,0	63,4	23,0	24,2	25,4	12,8	13,8	14,7	U	3
9604	20:37	60,7	62,1	63,4	23,0	24,2	25,4	12,7	13,7	14,7	U	3
9873	20:38	60,7	62,0	63,4	23,0	24,2	25,4	12,8	13,7	14,7	R	3
10106	20:39	60,7	62,1	63,4	22,9	24,1	25,3	12,9	13,8	14,8	M	3
10549	20:42	60,9	62,2	63,5	22,6	23,7	24,9	13,1	14,1	15,0	U	2
10824	20:43	60,9	62,2	63,5	22,3	23,4	24,6	13,5	14,4	15,3	U	2
11047	20:43	60,7	62,0	63,2	22,8	23,9	25,1	13,2	14,1	15,0	U	5
11351	20:45	60,9	62,2	63,5	22,9	24,0	25,1	12,9	13,8	14,7	U	3
11801	20:46	60,6	61,8	63,1	23,8	24,9	26,0	12,4	13,3	14,2	M	5
12155	20:46	60,6	61,8	63,1	23,8	24,9	26,0	12,4	13,3	14,1	M	4
12426	20:46	60,7	61,9	63,1	23,8	24,9	26,0	12,4	13,3	14,1	U	3
12868	20:47	60,5	61,7	62,9	23,5	24,6	25,6	12,8	13,7	14,5	U	2
13213	20:52	60,1	61,2	62,4	24,2	25,2	26,3	12,7	13,5	14,3	M	1
13659	20:53	59,9	61,1	62,3	24,0	25,1	26,1	13,0	13,8	14,6	U	2
13918	20:54	60,4	61,5	62,7	23,7	24,7	25,7	13,0	13,8	14,6	U	3
14188	20:55	60,4	61,6	62,7	23,7	24,7	25,7	12,9	13,8	14,6	U	1
14341	20:55	60,6	61,7	62,9	23,6	24,6	25,6	12,9	13,7	14,5	U	1
14685	20:56	60,6	61,7	62,8	23,6	24,6	25,6	12,9	13,7	14,4	U	1

Nº VOTOS	HORA	(FRI) (PRD) (OTROS)	TIPO	ESTRATO		
14864	20:56	60,6	61,7	62,8	23,6	24,6	25,6	12,9	13,7	14,5	U	1
15314	20:57	60,7	61,8	62,9	23,6	24,6	25,5	12,8	13,6	14,4	U	2
15574	21:02	61,3	62,4	63,5	23,2	24,1	25,1	12,7	13,5	14,3	U	3
15924	21:02	61,4	62,5	63,5	23,1	24,1	25,0	12,7	13,5	14,2	U	3
16204	21:03	61,4	62,4	63,5	23,1	24,0	25,0	12,8	13,6	14,3	U	2
16604	21:04	61,3	62,3	63,3	22,8	23,8	24,7	13,2	13,9	14,7	U	2
17173	21:05	60,5	61,5	62,5	22,6	23,5	24,4	14,3	15,0	15,8	U	2
17450	21:06	60,7	61,7	62,7	22,4	23,3	24,2	14,2	15,0	15,7	U	1
17648	21:06	60,5	61,5	62,5	22,9	23,8	24,6	14,0	14,8	15,5	M	1
17901	21:06	60,7	61,7	62,7	22,7	23,6	24,4	14,0	14,7	15,5	M	1
18181	21:07	60,8	61,8	62,8	22,9	23,8	24,6	13,7	14,5	15,2	M	3
18539	21:07	60,8	61,8	62,8	22,9	23,8	24,6	13,7	14,4	15,1	M	1
18768	21:08	60,8	61,8	62,8	22,9	23,8	24,6	13,7	14,4	15,1	U	1
19008	21:08	60,9	61,8	62,8	22,9	23,7	24,6	13,7	14,4	15,2	M	1
19315	21:09	60,7	61,7	62,7	23,2	24,1	24,9	13,5	14,2	14,9	U	1
19672	21:10	60,7	61,7	62,6	23,3	24,2	25,0	13,5	14,2	14,9	U	1
19907	21:11	60,6	61,6	62,6	23,4	24,2	25,0	13,5	14,2	14,9	U	3
20309	21:11	60,8	61,7	62,7	23,1	24,0	24,8	13,6	14,3	15,0	U	2
20602	21:11	61,0	61,9	62,9	23,0	23,9	24,7	13,6	14,2	14,9	U	2
20913	21:12	60,9	61,8	62,8	23,1	23,9	24,7	13,6	14,3	15,0	U	2
21246	21:13	61,0	62,0	62,9	23,0	23,8	24,6	13,5	14,2	14,9	U	2
21490	21:15	61,0	62,0	62,9	23,0	23,8	24,6	13,5	14,2	14,9	U	5
21778	21:15	61,0	62,0	62,9	23,0	23,8	24,6	13,6	14,2	14,9	M	3
22201	21:17	61,1	62,0	62,9	22,9	23,7	24,5	13,6	14,3	14,9	U	2
22335	21:17	61,1	62,0	62,9	23,2	24,0	24,7	13,4	14,0	14,7	U	1
22656	21:20	61,6	62,5	63,4	22,9	23,7	24,4	13,2	13,9	14,5	M	3
23032	21:20	61,7	62,6	63,4	22,9	23,7	24,5	13,1	13,8	14,4	U	1
23274	21:21	61,6	62,4	63,3	23,0	23,8	24,5	13,2	13,8	14,4	M	3
23713	21:22	61,2	62,1	62,9	23,7	24,5	25,2	12,9	13,5	14,1	M	2
24021	21:22	61,4	62,2	63,1	23,7	24,5	25,3	12,7	13,3	13,9	M	3
24488	21:22	61,3	62,2	63,1	23,8	24,6	25,3	12,7	13,3	13,9	U	1
24881	21:24	61,3	62,1	63,0	23,8	24,6	25,4	12,7	13,3	13,9	U	2
25144	21:26	61,4	62,3	63,1	23,8	24,6	25,3	12,6	13,2	13,8	U	1
25915	21:27	61,4	62,3	63,0	23,6	24,3	25,1	12,9	13,5	14,1	U	2
26178	21:29	61,0	61,9	62,7	24,1	24,9	25,6	12,7	13,3	13,9	M	2
26448	21:31	61,1	62,0	62,8	24,0	24,8	25,5	12,7	13,3	13,9	M	2
26787	21:42	61,4	62,2	63,0	23,8	24,6	25,3	12,7	13,3	13,8	M	4
27283	21:49	61,3	62,1	62,9	24,3	25,1	25,8	12,5	13,2	13,8	M	5
27550	22:11	60,6	61,4	62,2	24,7	25,4	26,2	12,6	13,3	13,7	R	2
27828	22:12	60,5	61,4	62,2	24,7	25,5	26,2	12,6	13,2	13,7	R	2
28122	22:13	60,5	61,3	62,1	25,0	25,7	26,4	12,5	13,1	13,6	M	1
28354	22:14	60,8	61,6	62,4	24,8	25,5	26,3	12,3	12,9	13,4	U	2
28591	22:15	60,7	61,6	62,4	24,8	25,6	26,3	12,3	12,9	13,4	M	1
28806	22:15	60,7	61,5	62,3	24,9	25,6	26,3	12,4	12,9	13,5	U	2
29219	22:17	60,7	61,5	62,2	24,9	25,6	26,3	12,4	12,9	13,5	U	2
29273	22:18	60,6	61,4	62,2	24,9	25,6	26,3	12,4	13,0	13,5	R	4
29690	22:18	60,7	61,4	62,2	25,1	25,8	26,5	12,2	12,7	13,3	U	3
29727	22:19	60,6	61,4	62,2	25,1	25,8	26,5	12,2	12,7	13,3	R	4
29992	22:20	60,6	61,4	62,2	25,3	26,0	26,7	12,1	12,6	13,2	M	3
30268	22:22	60,7	61,5	62,2	25,2	25,9	26,6	12,1	12,6	13,1	M	3
30534	22:31	60,7	61,5	62,2	25,1	25,8	26,5	12,2	12,7	13,2	U	2
30851	22:32	60,7	61,5	62,2	25,2	25,9	26,5	12,2	12,7	13,2	M	5
31127	22:44	60,9	61,7	62,4	24,9	25,6	26,3	12,2	12,7	13,2	M	3
31376	22:45	60,5	61,3	62,0	25,4	26,1	26,8	12,1	12,6	13,1	U	3
31645	22:45	60,4	61,2	62,0	25,5	26,2	26,8	12,1	12,6	13,2	M	3
31885	22:51	60,8	61,6	62,3	25,4	26,1	26,8	11,8	12,4	12,9	U	2
32252	22:52	61,3	62,0	62,8	25,1	25,8	26,5	11,7	12,2	12,7	U	2
32769	22:54	60,9	61,6	62,4	25,1	25,8	26,4	12,1	12,6	13,1	U	2
33174	23:03	60,9	61,7	62,4	25,1	25,7	26,4	12,1	12,6	13,1	U	2
33526	23:04	61,1	61,8	62,5	25,0	25,7	26,4	12,0	12,5	13,0	U	2

INTERVALOS DE CONFIANZA PARA LAS ESTIMACIONES PONDERADAS
DE LA DISTRIBUCION DEL VOTO EN ELECCIONES PARA PTE. MUNICIPAL
MÉRIDA, YUC.

Nº VOTOS	HORA	(PAN)	PRI)	PRD)	TIPO	ESTRATO		
368	18:33	35,2	42,4	49,6	48,2	55,4	62,6	-0,3	0,3	1,0	U	3
997	19:10	47,1	51,5	55,9	41,0	45,4	49,8	-0,1	0,7	1,4	U	5
1300	19:11	49,3	53,3	57,0	40,5	44,4	48,2	0,0	0,5	1,0	U	5
1541	19:12	47,2	50,8	54,3	43,4	46,9	50,5	0,0	0,5	0,9	R	1
1846	19:15	46,9	50,1	53,3	44,2	47,5	50,7	0,0	0,5	0,9	U	5
2219	19:17	47,2	50,1	53,1	44,8	47,8	50,8	0,0	0,4	0,7	U	2
2560	19:25	45,4	48,2	50,9	47,1	49,9	52,6	0,0	0,4	0,7	R	1
2846	19:27	45,2	47,8	50,4	47,3	49,9	52,5	0,1	0,4	0,8	U	5
3175	19:28	44,7	47,1	49,6	48,0	50,5	53,0	0,2	0,5	0,9	U	4
3443	19:29	44,8	47,1	49,5	48,1	50,5	52,9	0,2	0,5	0,8	U	5
3943	19:32	45,3	47,6	49,8	47,8	50,0	52,2	0,2	0,5	0,7	U	4
4200	19:33	45,8	48,0	50,1	47,5	49,7	51,8	0,1	0,4	0,7	U	5
4621	19:33	46,3	48,3	50,4	47,4	49,4	51,5	0,1	0,4	0,6	U	5
5036	19:34	46,4	48,4	50,3	47,4	49,4	51,3	0,1	0,4	0,6	U	5
5575	19:36	47,1	49,0	50,8	47,0	48,9	50,7	0,2	0,4	0,6	U	5
6052	19:40	47,2	49,0	50,8	47,1	48,9	50,6	0,2	0,4	0,6	U	4
6498	19:42	48,3	50,0	51,8	46,0	47,7	49,4	0,2	0,4	0,6	U	3
7071	19:44	47,7	49,3	51,0	46,8	48,4	50,1	0,2	0,4	0,6	U	2
7545	19:44	47,0	48,6	50,2	47,7	49,3	50,9	0,2	0,4	0,6	U	3
8173	19:49	47,6	49,1	50,7	47,3	48,8	50,4	0,2	0,4	0,6	U	3
8577	19:52	47,1	48,6	50,1	47,9	49,4	50,9	0,2	0,4	0,6	U	4
8849	19:52	47,4	48,9	50,3	47,7	49,2	50,6	0,2	0,4	0,6	U	4
9156	19:55	46,8	48,3	49,7	48,2	49,7	51,1	0,2	0,4	0,6	U	1
9453	19:57	47,1	48,6	50,0	48,0	49,5	50,9	0,2	0,4	0,5	U	4
9803	19:57	47,1	48,5	49,9	48,2	49,6	51,0	0,2	0,4	0,5	U	4
10139	19:58	46,6	48,0	49,4	48,7	50,1	51,5	0,2	0,4	0,5	U	5
10504	19:59	46,6	48,0	49,3	48,7	50,1	51,5	0,2	0,4	0,5	U	4
10805	20:00	46,6	47,9	49,2	48,8	50,1	51,5	0,2	0,4	0,5	U	4
11160	20:02	47,4	48,7	50,0	48,0	49,4	50,7	0,2	0,4	0,5	U	3
11512	20:03	47,4	48,7	50,0	48,0	49,3	50,6	0,2	0,4	0,5	U	4
11910	20:05	47,5	48,7	50,0	48,0	49,3	50,6	0,2	0,4	0,5	U	5
12312	20:07	47,5	48,7	50,0	48,1	49,3	50,6	0,2	0,4	0,5	U	4
12703	20:08	47,9	49,2	50,4	47,6	48,9	50,1	0,2	0,4	0,5	U	3
13096	20:09	47,7	48,9	50,1	47,8	49,1	50,3	0,2	0,3	0,5	U	5
13377	20:13	47,7	48,9	50,1	47,8	49,0	50,2	0,2	0,3	0,5	U	5
13769	20:14	47,7	48,9	50,1	47,9	49,1	50,3	0,2	0,3	0,5	U	3
14168	20:14	47,9	49,0	50,2	47,8	49,0	50,2	0,2	0,3	0,5	U	5
14521	20:20	47,7	48,9	50,0	48,0	49,2	50,3	0,2	0,3	0,5	U	5
14867	20:22	47,8	49,0	50,1	48,0	49,1	50,2	0,2	0,3	0,5	U	4
15320	20:23	48,0	49,1	50,2	47,9	49,0	50,2	0,2	0,3	0,5	U	4
15685	20:24	48,3	49,4	50,5	47,7	48,8	49,9	0,2	0,3	0,5	U	3
15867	20:27	47,7	48,8	49,9	48,2	49,3	50,4	0,2	0,4	0,5	U	2
16314	20:27	47,7	48,8	49,9	48,2	49,3	50,4	0,2	0,4	0,5	U	3
16654	20:29	47,8	48,8	49,9	48,2	49,3	50,3	0,2	0,4	0,5	U	4
17194	20:30	47,9	49,0	50,0	48,0	49,1	50,2	0,2	0,4	0,5	U	3
17784	20:30	48,1	49,1	50,2	47,9	48,9	50,0	0,3	0,4	0,5	U	4
18043	20:33	48,1	49,1	50,1	47,9	49,0	50,0	0,3	0,4	0,5	R	1
18449	20:34	47,8	48,8	49,8	48,3	49,3	50,3	0,3	0,4	0,5	U	2
18712	20:34	48,1	49,1	50,1	48,0	49,0	50,0	0,2	0,4	0,5	U	3
19014	20:36	48,1	49,1	50,1	48,0	49,0	50,0	0,2	0,4	0,5	U	3
19545	20:37	47,9	48,9	49,9	48,2	49,2	50,2	0,3	0,4	0,5	U	5
20017	20:39	48,2	49,2	50,1	48,0	49,0	49,9	0,3	0,4	0,5	U	5
20415	20:42	48,2	49,2	50,1	48,0	49,0	49,9	0,3	0,4	0,5	U	4
20797	20:42	48,4	49,3	50,3	47,9	48,8	49,8	0,3	0,4	0,5	U	3
21169	20:43	48,4	49,3	50,3	47,9	48,8	49,8	0,3	0,4	0,5	U	3
21582	20:44	48,5	49,5	50,4	47,7	48,7	49,6	0,3	0,4	0,5	U	4
22139	20:47	48,5	49,4	50,4	47,8	48,7	49,6	0,3	0,4	0,5	U	5
22610	20:47	48,6	49,5	50,4	47,7	48,6	49,6	0,3	0,4	0,5	U	5

ANEXO B

* VOTOS	HORA	(PAN) (FRI) (PRD)	TIPO	ESTRATO		
22904	20:49	48.4	49.3	50.2	48.0	48.9	49.8	0.3	0.4	0.5	U	5
23119	20:49	48.4	49.3	50.2	48.0	48.9	49.8	0.3	0.4	0.5	U	4
23655	20:52	48.4	49.3	50.2	48.0	48.9	49.8	0.3	0.4	0.5	U	3
23944	20:53	48.4	49.3	50.2	48.0	48.9	49.8	0.3	0.4	0.5	U	5
24374	20:53	48.5	49.4	50.3	47.9	48.8	49.7	0.3	0.4	0.5	U	4
24723	20:54	48.5	49.4	50.3	47.9	48.8	49.7	0.3	0.4	0.5	U	5
25210	20:54	48.9	49.7	50.6	47.6	48.5	49.4	0.3	0.4	0.5	U	3
25693	20:55	49.0	49.9	50.7	47.5	48.3	49.2	0.3	0.4	0.5	U	4
26191	20:56	48.9	49.8	50.7	47.6	48.4	49.3	0.3	0.4	0.5	U	4
26698	20:57	49.5	50.3	51.2	47.0	47.9	48.8	0.3	0.4	0.5	U	3
27039	20:59	49.6	50.4	51.3	47.0	47.8	48.7	0.3	0.4	0.5	U	2
27339	21:00	49.6	50.4	51.3	47.0	47.8	48.7	0.3	0.4	0.4	U	4
27732	21:01	49.7	50.5	51.4	46.9	47.7	48.5	0.3	0.4	0.5	U	4
28091	21:08	49.6	50.5	51.3	46.9	47.8	48.6	0.3	0.4	0.4	U	4
28477	21:10	49.7	50.5	51.3	46.9	47.8	48.6	0.3	0.4	0.4	U	5
28892	21:11	49.7	50.5	51.3	46.9	47.8	48.6	0.3	0.4	0.4	U	4
29313	21:12	49.9	50.7	51.6	46.7	47.5	48.3	0.3	0.4	0.4	U	5
29801	21:12	49.9	50.7	51.5	46.7	47.5	48.3	0.3	0.4	0.5	U	4
30163	21:12	49.9	50.7	51.5	46.7	47.5	48.3	0.3	0.4	0.5	U	5
30520	21:16	49.6	50.4	51.2	47.0	47.8	48.6	0.3	0.4	0.5	U	4
30919	21:19	49.7	50.5	51.3	47.0	47.8	48.6	0.3	0.4	0.5	U	4
31233	21:27	49.4	50.3	51.0	47.2	48.0	48.8	0.3	0.4	0.5	U	4
31609	21:30	49.3	50.1	50.9	47.4	48.1	48.9	0.3	0.4	0.5	U	4
32278	21:31	49.0	49.8	50.5	47.8	48.5	49.3	0.3	0.4	0.5	U	4
32912	21:32	49.0	49.8	50.5	47.7	48.5	49.3	0.3	0.4	0.5	U	3
33229	21:34	49.1	49.9	50.6	47.6	48.4	49.1	0.3	0.4	0.5	U	4
33674	21:36	49.0	49.8	50.6	47.7	48.5	49.2	0.3	0.4	0.5	U	3
34191	21:37	49.1	49.8	50.6	47.6	48.4	49.1	0.3	0.4	0.5	U	4
34652	22:00	49.2	50.0	50.7	47.5	48.2	49.0	0.3	0.4	0.5	U	4
35139	22:01	49.2	49.9	50.7	47.5	48.3	49.0	0.3	0.4	0.5	U	5
35678	22:41	49.2	49.9	50.7	47.5	48.3	49.0	0.3	0.4	0.5	U	4

**INTERVALOS DE CONFIANZA PARA LAS ESTIMACIONES PONDERADAS
DE LA DISTRIBUCION DEL VOTO EN ELECCIONES PARA PTE. MUNICIPAL
OCOYOACAC, EDO. DE MEXICO**

Nº VOTOS (PRI) (PRD) (OTROS)	TIPO	ESTRATO		
365	27.3	34.3	41.2	14.2	20.0	25.8	38.5	45.0	53.0	U	1
734	26.7	31.5	36.2	13.2	17.0	20.9	46.4	51.5	56.6	U	1
979	31.4	35.7	39.9	26.8	31.0	35.1	29.2	33.4	37.6	R	1
1347	27.3	30.8	34.3	21.5	24.8	28.1	40.6	44.4	48.2	R	1
1768	30.5	33.7	36.8	30.6	33.7	36.9	29.5	32.6	35.7	R	2
2125	29.5	32.3	35.1	29.6	32.4	35.2	32.4	35.3	38.2	R	2
2434	29.6	32.2	34.8	30.6	33.3	35.9	31.8	34.5	37.2	U	2
2728	30.3	32.8	35.3	30.9	33.4	35.9	31.3	33.8	36.3	U	2
3128	27.1	29.4	31.7	28.0	30.3	32.6	37.8	40.3	42.7	U	2
3437	28.1	30.3	32.4	28.8	31.0	33.2	36.4	38.7	41.0	U	2
3769	28.0	30.1	32.2	27.1	29.2	31.3	38.5	40.7	42.9	U	1
4114	28.1	30.1	32.1	27.3	29.3	31.3	38.4	40.6	42.7	U	1
4289	28.1	30.1	32.0	26.2	28.1	30.0	39.7	41.8	43.9	R	2
4453	28.3	30.2	32.1	26.4	28.3	30.1	39.5	41.5	43.6	R	2
4719	32.9	34.9	36.8	23.8	25.5	27.3	37.6	39.6	41.6	R	1
5306	36.9	38.7	40.6	21.8	23.5	25.1	36.0	37.8	39.7	R	1
5725	36.2	38.0	39.8	22.1	23.7	25.2	36.6	38.4	40.1	R	1
6017	35.2	37.0	38.7	25.3	26.9	28.5	34.4	36.2	37.9	U	1
6232	35.3	37.0	38.7	25.3	26.8	28.4	34.4	36.1	37.8	U	1
6470	36.1	37.8	39.5	25.9	27.5	29.0	33.1	34.7	36.4	U	1
6706	35.8	37.5	39.1	26.3	27.8	29.3	33.1	34.7	36.3	U	1

DIFERENCIA DE PROPORCIONES

Si se tienen dos poblaciones con proporciones P_1 y P_2 con varianzas conocidas, para probar la hipótesis nula $(p_1 - p_2) = 0$ frente a la hipótesis alternativa $(p_1 - p_2) \neq 0$, entonces el estadístico de la prueba está dado por

$$Z = \frac{(P_1 - P_2) - (p_1 - p_2)}{\sqrt{\left(\frac{p_1 q_1}{n_1}\right) + \left(\frac{p_2 q_2}{n_2}\right)}}$$

Para un nivel de confianza de 95 %, se rechaza la hipótesis nula de que las dos proporciones son iguales cuando $|z| > 1.96$, siempre y cuando p_1 y p_2 tienen una distribución muestral normal y n_1 y n_2 son resultado de selecciones de muestras independientes y grandes (digamos $n > 30$).

PRUEBA JI-CUADRADA

Cuando se tiene una tabla de contingencia de una muestra multinomial, la hipótesis nula de independencia estadística es $H_0: \pi_{ij} = \pi_{i.} \pi_{.j}$ para toda i, j . Para probar H_0 se puede utilizar el estadístico Pearson χ^2 .

$$\chi^2 = \sum \sum \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

donde

n_{ij} valores observados en la celda ij

$\hat{m}_{ij} = n p_{i.} p_{.j}$ estimación de los valores esperados

Este estadístico de prueba debe compararse con el valor de una χ^2 con $(I-1)(J-1)$ grados de libertad a determinado nivel de confianza. Específicamente a un 95 % de confianza y con $(4-1)(2-1) = 3$ grados de libertad $\chi^2 = 0.3518$.

MORELOS

Diferencia de proporciones

	Todos los estratos	Estrato 1	Estrato 2	Estrato 3
PAN	9.5	13.5	5.4	10.2
PRI	61.8	57.7	64.5	62.1
PRD	25.7	25.9	27.3	24.5
OTROS	3.0	2.8	2.8	3.2
N° votantes	33526	16195	19534	31323
		Diferencia entre T y 1	Diferencia entre T y 2	Diferencia entre T y 3
PAN		-12.8	18.0	-3.0
PRI		8.7	-6.2	-0.8
PRD		-0.5	-4.0	3.5
OTROS		1.3	1.3	-1.5
			Diferencia entre 1 y 2	Diferencia entre 1 y 3
PAN			25.8	10.4
PRI			-13.1	-9.3
PRD			-3.0	3.3
OTROS			0.0	-2.4
				Diferencia entre 2 y 3
PAN				-20.4
PRI				5.5
PRD				7.0
OTROS				-2.6

Prueba ji-cuadrada

	χ^2
Todos los estratos y estrato 1	181.1072
Todos los estratos y estrato 2	420.1810
Todos los estratos y estrato 3	388.7090
Estrato 1 y estrato 2	812.2863
Estrato 1 y estrato 3	317.6829
Estrato 2 y estrato 3	582.2426

MERIDA

Diferencia de proporciones

	Todos los estratos	Estrato 1	Estrato 2	Estrato 3
PAN	49.9	51.1	47.6	51.9
PRI	48.3	47.1	50.6	46.2
PRD	0.4	0.4	0.4	0.5
OTROS	1.4	1.4	1.4	1.4
N° votantes	35678	32655	27283	11418
		Diferencia entre T y 1	Diferencia entre T y 2	Diferencia entre T y 3
PAN		-3.1	5.7	-3.7
PRI		3.1	-5.7	3.9
PRD		0.0	0.0	-1.4
OTROS		0.0	0.0	-0.2
			Diferencia entre 1 y 2	Diferencia entre 1 y 3
PAN			8.5	-1.5
PRI			-8.5	1.7
PRD			0.0	-1.3
OTROS			0.0	-0.2
				Diferencia entre 2 y 3
PAN				-7.7
PRI				7.9
PRD				-1.3
OTROS				-0.2

Prueba ji-cuadrada

	χ^2
Todos los estratos y estrato 1	208.6834
Todos los estratos y estrato 2	208.1458
Todos los estratos y estrato 3	4.7627
Estrato 1 y estrato 2	427.6100
Estrato 1 y estrato 3	175.8732
Estrato 2 y estrato 3	229.5778

OCOYOACAC**Diferencia de proporciones**

	Todos los estratos	Estrato 1	Estrato 2
PAN	9.5	12.0	8.1
PRI	37.5	30.2	41.4
PRD	27.8	32.3	25.4
OTROS	25.2	25.5	25.2
Nº votantes	6706	2429	4277

	Diferencia entre T y 1	Diferencia entre T y 2
PAN	-3.3	2.5
PRI	6.6	-4.1
PRD	-4.1	2.8
OTROS	-0.3	0.0

	Diferencia entre 1 y 2
PAN	5.0
PRI	-9.3
PRD	6.0
OTROS	0.3

Prueba ji-cuadrada

	χ^2
Todos los estratos y estrato 1	21.4873
Todos los estratos y estrato 2	50.2027
Estrato 1 y estrato 2	103.3279

Los resultados de los cálculos referentes a la diferencia entre áreas urbanas y rurales se presentan a continuación.

Diferencia de proporciones

	MORELOS		OCOYOACAC	
	URBANA	RURAL	URBANA	RURAL
PAN	13.5	4.3	8.9	10.1
PRI	62.0	61.5	30.6	45.5
PRD	20.8	32.1	31.4	23.6
OTROS	3.6	2.1	29.1	20.8
N° votantes	19,939	13,587	3,704	3,002

	MORELOS	OCOYOACAC
PAN	30.9	-1.7
PRI	0.9	-12.6
PRD	-22.9	7.2
OTROS	8.3	7.9

Prueba ji-cuadrada

	χ^2
Morelos: urbana y rural	1167.3520
Ocoyoacac: urbana y rural	182.4478

CUESTIONARIO DE LA ENCUESTA POSVOTO EN YUCATAN

En la EP realizada el 28 de noviembre de 1993 en el estado de Yucatán se utilizaron 2 cuestionarios: uno para el municipio de Mérida y otro para los demás municipios que conforman el estado de Yucatán. Cada una de estas dos versiones se dividió en 3 tipos de cuestionarios que se diferencian en la rotación de opciones en algunas preguntas.

Cuestionario para el Estado de Yucatán:

A. Tipo de cuestionario

B. Perdona, ¿pudo usted votar?

(1) Sí

(2) No (SI NO PUDO: SEGUIR EL CUESTIONARIO HASTA LA PREGUNTA (E) Y SUSPENDER)

C. Sexo (ENCUESTADOR: ANOTE SIN PREGUNTAR)

D. ¿Cuántos años cumplidos tiene usted?

E. ¿Hasta qué año estudió usted?

(1) Ninguno

(2) Primaria (completa o incompleta)

(3) Secundaria o Preparatoria (completa o incompleta)

(4) Universidad y más (completa o incompleta)

G. En general, ¿está usted de acuerdo o en desacuerdo con la manera como ha trabajado la actual gobernadora Dulce María Sauri Riancho?

(1) De acuerdo

(2) En parte (sólo por respuesta espontánea)

(3) En desacuerdo

(4) No conozco la forma en que trabajó (espontánea)

H. En general, ¿está usted de acuerdo o en desacuerdo con la manera como está gobernando el Presidente Salinas?

- (1) Acuerdo
- (2) Desacuerdo
- (3) No sabe - no contestó

I. Comparada con la situación económica que tenía Yucatán hace un año, ¿cómo diría usted que es la situación actual de su estado, mejor o peor?

- (1) Mejor
- (2) Igual (sólo espontánea)
- (3) Peor
- (4) No sabe - no contestó

J. Sumando lo que ganan todos en su casa, ¿cuál es su ingreso familiar mensual?
(ENCUESTADOR: SI EL ENTREVISTADO DUDA AL CONTESTAR, AYUDELO LEYENDOLE LOS DIFERENTES RANGOS DE INGRESO)

- (1) entre 0 - 1 sal. mínimo N\$ (0 a 428)
- (2) entre 1 - 3 sal. mínimos N\$ (429 a 1,284)
- (3) entre 3 - 5 sal. mínimos N\$ (1,285 a 2,140)
- (4) entre 5 - 7 sal. mínimos N\$ (2,141 a 2,995)
- (5) entre 7 - 10 sal. mínimos N\$ (2,996 a 4,279)
- (6) 10 a más sal. mínimos N\$ (4,280 y más)
- (7) No sabe - no contestó

K. ¿Cuál es su ocupación principal en este momento?

- (1) Estoy desempleado y estoy buscando trabajo.
- (2) Trabajador de tiempo completo en el gobierno (o paraestatal).
- (3) Trabajador de tiempo completo en el sector privado (no gobierno).
- (4) Profesionista independiente.
- (5) Trabajador por cuenta propia (taxista, vendedor ambulante, etc.).
- (6) Trabajador de tiempo parcial solamente.
- (7) Estudiante no trabaja.

(8) Ama de casa (no trabaja)

(9) Jubilado (no trabaja)

(10) Otro

L. ¿Cuándo decidió votar por el candidato o partido por el que votó hoy? (NO LEA OPCIONES)

(1) Siempre vota por el mismo partido: Hace más de 4 años

(2) Siempre vota por el mismo partido: Desde hace 4 años

(3) Desde que supo quiénes eran los candidatos

(4) Durante el último mes

(5) Entre ayer y hoy (en los últimos días)

(6) Otra:

(7) No sabe - no contestó

M. ¿De cuál de los siguientes medios recibió usted la información que más le ayudó para decidir por quién votar? (ENCUESTADOR LEA OPCIONES 1 A 4)

(1) Lo escuchó directamente (mitines, o visitas domiciliarias)

(2) Por la propaganda en la calle (bardas, carteles, volantes)

(3) Por la radio o la televisión

(4) Por el periódico

(5) Ninguna de las anteriores

(6) No sabe - no contestó

N. ¿Qué es lo primero que debe realizar el próximo gobierno? (ENCUESTADOR: LEA OPCIONES (1) Y (2))

(1) Reforzar el derecho a elegir libremente a los gobernantes

(2) Continuar con el rumbo actual del país

(3) Ninguna de las anteriores (sólo espontánea)

(4) No sabe

(5) No contestó

O. ¿Por cuál de las siguientes dos razones vino usted a votar? (ENCUESTADOR: LEA OPCIONES (1) Y (2))

- (1) Por la estabilidad y la paz social
- (2) Por el cambio
- (3) Ninguna de las anteriores (sólo espontánea)
- (4) No sabe - no contestó

P. Y ¿por cuál de las siguientes dos razones vino usted hoy a votar? (ENCUESTADOR: LEA OPCIONES (1) Y (2))

- (1) Porque es una obligación
- (2) Para manifestar su desacuerdo
- (3) Ninguna de las anteriores (sólo espontánea)
- (4) No sabe - no contestó

Q. ¿Qué periódico lee usted con más frecuencia? (ENCUESTADOR: SOLO LEA LAS TRES PRIMERAS OPCIONES)

- (1) El Diario de Yucatán
- (2) Novedades de Yucatán
- (3) Por Esto
- (4) Otro
- (5) Ninguno
- (6) No sabe - no contestó.
- (7) No leo el periódico

ENCUESTADOR: ENTREGUE AL ENTREVISTADO EL CUESTIONARIO, AL HACERLO MENCIONE EL NUMERO DE PREGUNTAS QUE TIENE QUE RESPONDER Y EN QUE CONSISTE CADA UNA DE ELLAS.

MENCIONE QUE EL AUTOLLENADO PERMITE MANTENER LA CONFIDENCIALIDAD DE LA INFORMACION Y RESPETAR EL SECRETO DE SU VOTO.

INDIQUE QUE PARA CADA PREGUNTA DEBE CIRCULAR EL NUMERO CORRESPONDIENTE.

R. ¿Por cuál partido o candidato votó para gobernador (marque el partido por el que votó)

- (1) PAN Ana Rosa Payán
- (2) PRI Federico Granja Ricalde
- (3) PPS José Zacarías Alejos
- (4) PRD Enrique Montalvo Ortega
- (5) PFCRN Ma. Amalia R. Martínez
- (6) PARM Francisco Luna Martín
- (7) PDM Juan Bojórquez Flores
- (8) PT Enrique Montalvo Ortega
- (9) PVEM Hermelinda Padilla Can
- (10) Otro candidato o partido: (especifique)

S. ¿Por qué votó por este candidato o partido? (circule sólo la opción que más se parezca a lo que usted opina)

- (1) Simpatizo con el candidato
- (2) Porque tiene experiencia (para gobernar)
- (3) Ayudará a que mejore mi situación
- (4) Otro: (especifique)

Preguntas adicionales en el cuestionario aplicado en Mérida:

F. En general, ¿está usted de acuerdo o en desacuerdo con la manera como trabajó la presidenta municipal Ana Rosa Payán?

- (1) De acuerdo
- (2) En parte (sólo por respuesta espontánea)
- (3) En desacuerdo
- (4) No conozco la forma en que trabajó (espontánea)
- (5) No sabe - no contestó

T. ¿Por cuál partido o candidato votó para presidente municipal? (Marque el partido por el que votó)

- (1) PAN Luis Correa Mena
- (2) PRI Orlando Paredes
- (3) PPS Jaime Domínguez Pech
- (4) PRD Rubén Darío Domínguez
- (6) PARM Teresa Pallares Rivero
- (7) PDM José Humberto Barrera Castillo
- (8) PT Carlos Bojórquez U.
- (9) PVEM Lucía Bacab Sulub
- (10) Otro candidato o partido: (especifique)

U. ¿Por qué votó por este candidato o partido? (circule sólo la opción que más se parezca a lo que usted opina)

- (1) Porque tiene experiencia (para gobernar)
- (2) Simpatizo con el candidato
- (3) Ayudará a que mejore mi situación
- (4) Otro: (especifique)

V. ¿Votó usted por el mismo partido en las elecciones de 1990 para elegir presidente municipal de Mérida? (Circule sólo una opción)

- (1) Sí
- (2) No
- (3) No recuerdo
- (4) No voté