12

2 e j

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ALGUNOS ASPECTOS ESTADISTICOS EN CONTEOS RAPIDOS.

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

IGNACIO HIRAM BENITEZ ORTIZ

DIRECTORA: DRA. GUILLERMINA ESLAVA GOMEZ.

1 9 9 8

258722.

VNIVERSDAD NACIONAL
AVFNºMA DE
MEXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:

ALGUNOS ASPECTOS ESTADISTICOS EN CONTEOS RAPIDOS.

realizado por Ignacio Hiram Benítez Ortiz.

con número de cuenta 8152774-8 , pasante de la carrera de ACTUARIO

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dra. Guillermina Eslava Gómez.
Propietario	Dr. Ignacio Méndez Ramírez.
Propietario	Mat. Margarita Elvira Chávez Cano.
Suplente	M. en C. Salvador Zamora Muñoz.
Suplente	M. en C. Leticia G[illegible]no Valdelamar.

Consejo Departamental de Matemáticas

Mtra. en A.P. María del Pilar Alonso Reyes

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMATICAS

A mis padres
José Benítez Romero
Enedina Ortiz Aguiñaga
por su apoyo incondicional y confianza

A Rosa María Navarro Vega
por su infinito amor y ternura

A Stephanie, María Fernanda y Montserrat
por su cariño y entusiasmo por vivir

A los amigos y compañeros de siempre

A la Dra. Guillermina Eslava Gómez por ser una personalidad importante en Estadística Aplicada y en reconocimiento por su decidido impulso y generoso apoyo.

Gracias a

Dr. Ignacio Méndez Ramírez

Mat. Margarita Elvira Chávez Cano

M. en C. Salvador Zamora Muñoz

M. en C. Leticia Gracia Medrano Valdelamar

investigadores y catedráticos en estadística, por sus comentarios al revisar este trabajo.

ÍNDICE

PREFACIO

Las encuestas electorales han adquirido presencia en México recientemente, son instrumentos cada vez más utilizados para medir la opinión del electorado previa a las elecciones. La evolución de las encuestas a partir de su introducción y utilización en la cultura política del país ya es palpable, esto se basa en los siguientes hechos; primero, se ha detectado reducción en la diferencia entre los pronósticos estadísticos y los resultados de las votaciones; segundo, los ciudadanos aceptan las encuestas con menos prejuicios y se resisten cada vez menos a contestarlas o permiten su realización en forma más amplia y libre, excepto en ciertas zonas rurales en el interior del país.

La controversia que rodea los resultados de la elecciones es un tema común que se discute en todos los niveles de la sociedad en la actualidad. El comparar ejercicios de estimaciones preliminares para conocer tendencias antes de los resultados oficiales, ha motivado el interés por investigar métodos alternativos para la realización de estos sondeos.

De los diversos métodos empleados para estimar resultados de votaciones en forma anticipada, el *conteo rápido* es uno de los utilizados con más frecuencia por su relativa facilidad y resultados aceptables. Consiste en recabar la información de la votación de una muestra de secciones al momento de cerrarse estas oficialmente, el día de las elecciones.

El objetivo del presente trabajo consiste en generar resultados por medio de simulación de conteos rápidos, a través de diversos procedimientos de selección de una muestra aleatoria de secciones electorales.

El muestreo aplicado a ejercicios de índole electoral ha probado ser, (en la mayoría de los casos) tan eficiente como el aplicado en investigaciones de mercado y niveles de audiencia. Esto es significativo porque la opinión en elecciones y en general la opinión pública, es muy inestable y difícil de medir.

El poder generar información en forma oportuna con el respaldo de la encuestas por muestreo, permite tomar decisiones en el diseño y desarrollo de cualquier programa de trabajo. Incluso se ha mencionado (Alberto Aziz Nassif, Dr. en sociología política miembro del Centro de Investigaciones y Estudios Superiores en Sociología) que no hay político que gobierne sin encuestas, ni partido político que plantee sus estrategias sin apoyo de estudios de opinión.

Las encuestas de opinión arrojan estimaciones del comportamiento de preferencias electorales, aunque algunas empresas no dan a conocer su metodología porque la consideran información confidencial.

Es importante que las encuestas se efectúen con rigor cívico, científico y ético al definir y aplicar los métodos estadísticos. Parte del público podrá evaluarlas no tan sólo por sus resultados, sino también por el diseño de muestra empleado empleado, el tamaño de muestra y la fecha de la realización además de otros detalles. Esto es vital sobre todo para la confiabilidad de las empresas encuestadoras.

Los propósitos que persigue el presente trabajo son: observar la tendencia en el resultado de simulaciones de conteos rápidos al estimar la proporción de votos en elecciones a gobernador en tres estados de la República Mexicana, así como obtener resultados del método de estimación de varianza llamado grupos aleatorios, al aplicarlo a las muestras generadas.

Con los resultados obtenidos, se pretende mostrar la aplicación estadística y computacional para estimar la preferencia electoral y coadyuvar a obtener mayor confianza y certeza en los resultados de este proceso de participación ciudadana. Se muestra también un método de estimación de varianza que por ser de fácil aplicación, permite comparar sus resultados con las fórmulas de varianza habitualmente utilizadas .

RESUMEN

Este trabajo se encuentra dividido en seis capítulos y un anexo. Describe la teoría de algunos diseños de muestreo, la estimación de la varianza de estimadores mediante grupos aleatorios y los resultados de simular conteos rápidos con diferentes diseños de muestreo y con tamaños de muestra fijos. Se obtienen estimaciones de la proporción de votos y de la varianza de los estimadores con los datos de las elecciones a gobernador en tres estados de la República Mexicana; en el anexo se presentan los programas fuentes.

En el capítulo 1 se presentan los conceptos relacionados con la teoría del muestreo y algunos de los diferentes diseños muestrales más utilizados: el muestreo aleatorio simple, muestreo sistemático, muestreo de conglomerados y el muestreo estratificado.

En el capítulo 2 se presenta un método de estimación de varianza llamado grupos aleatorios. Contempla dos casos : grupos aleatorios independientes y grupos aleatorios dependientes, además de su aplicación para cada diseño muestral.

En el capítulo 3 se definen el concepto de conteo rápido y el procedimiento de simulación aplicado a diferentes diseños de muestreo. Se presentan los métodos para calcular los estimadores de la proporción de votos y los estimadores de sus varianzas mediante grupos aleatorios independientes y dependientes .

En el capítulo 4 se reportan las tablas con los resultados numéricos de las simulaciones realizadas. La información que se presenta incluye una comparación de los resultados poblacionales con los obtenidos bajo diferentes diseños muestrales, se utiliza la información de las elecciones a gobernador efectuadas en 1995 en los estados de Guanajuato, Yucatán y Michoacán. Aunque de este último se tiene información parcial de 13 distritos de un total de 18.

En el capítulo 5 se presentan, para los tres estados, análiis de correlación lineal entre el número de electores en la lista nominal de secciones y los votos válidos registrados, así como un cuadro comparativo de los resultados de conteos rápidos extraoficiales con los presentados en forma oficial.

Los comentarios y conclusiones que se presentan en el capítulo 6, surgen de la comparación de los resultados de las simulaciones obtenidos bajo los diferentes diseños muestrales. Se comenta el desarrollo de los programas y se presentan las conclusiones. Para finalizar, se agregan algunas reflexiones sobre los alcances y limitaciones del presente análisis.

En el Anexo se presentan los programas fuente escritos en CLIPPER, (un lenguaje de programación de alto nivel basado en el lenguaje C) que se utilizan para generar estimadores de la proporción de votos y de su varianza.

Con los programas 1 y 2 se obtienen muestras aleatorias sin y con reemplazo respectivamente, utilizan una rutina recursiva que genera números aleatorios con repetición, y se aplican en el programa 3 para generar muestras aleatorias simples sin reemplazo.

La estimación de varianza se calcula sólo para los estimadores que utilizan exclusivamente información de la muestra.

Los programas 4 y 5 generan (en un muestreo aleatorio simple) 2 y 4 grupos aleatorios independientes respectivamente, y calculan los estimadores de varianza. Los programas 6 y 7 hacen lo propio para grupos aleatorios dependientes.

Para calcular los estimadores en un muestreo sistemático se utiliza el programa 8, la estimación de su varianza para 2 y 4 grupos aleatorios independientes y dependientes se obtiene de aplicar los programas 4,5,6 y 7, respectivamente.

Para el muestreo estratificado por distrito electoral, los estimadores se calculan mediante el programa 9; con los programas 10 y 11 se obtienen estimaciones de su varianza con grupos aleatorios independientes, y los programas 12 y 13 la estiman para grupos aleatorios dependientes.

Con el programa 14 se estiman las proporciones de votación para un muestreo estratificado urbano/ rural por casillas y sistemático con arranque aleatorio independiente en cada estrato, el programa 15 es un programa de filtrado de la información.

CAPÍTULO 1. ALGUNOS DISEÑOS DE MUESTREO

1.1 Introducción

La información de una gran variedad de características en una población es requerida para análisis políticos, de investigación de mercado, y por los responsables de planear programas económicos. Por cuestiones de tiempo y costo, la información se obtiene por medio de encuestas por muestreo. Necesitándose recurrir a diferentes métodos para seleccionar un subconjunto (una muestra) de todas las personas que poseen las características de interés (la población de sujetos), y observar las respuestas de los individuos que se entrevistan.

Una encuesta por muestreo es un estudio que involucra un subconjunto (o muestra) de individuos seleccionados de una población y sus mediciones observadas. Las medidas pueden utilizarse para estimar promedios, proporciones y totales con las que se pueden hacen inferencias de los valores o parámetros de la población. La validez o precisión de las estimaciones dependen de qué tan bueno sea el método de escoger la muestra y de qué tan bien se hicieron las mediciones, entre otros aspectos. Si el diseño fue aplicado correctamente, es posible estimar el margen de error para apreciar la exactitud de los resultados.

Algunas de las ventajas de una encuesta comparadas con la enumeración total o censo, consisten en que reducen el costo y los resultados se obtienen en menor tiempo y pueden ser más precisos, ya que en un censo puede haber más errores de medición por poca calidad del proceso de levantamiento de la información.

A continuación se describen conceptos relacionados con la terminología que se utiliza en el contexto del muestreo.

Unidad de muestreo : unidades como personas, animales, objetos o conjuntos de ellas, de las que se desean conocer características, opiniones, etc., se les denominará sujetos.

Población $\mathcal{P}$: Conjunto de N unidades identificables y diferentes llamadas unidades de muestreo.

Muestra: Subconjunto de cardinalidad n de la población, en el que sus elementos se escogen para medir una característica. La toma de una muestra de la población consiste de dos partes :

a) Escoger al azar un subconjunto de la población con probabilidades conocidas de selección para cada unidad o pueden escogerse de alguna otra manera y

b) Medir o evaluar la característica de interés en ese subconjunto y obtener los valores numéricos de la muestra.

Diseño de muestreo : Es el método de seleccionar una muestra de la población bajo estudio

Tipos de muestreo :

Muestreo errático o circunstancial: No se sigue ninguna regla para tomar la muestra, se efectúa por razones de comodidad, circunstancias o el juicio personal determinan que unidades deben incluirse. Se obtiene una parte de los valores de la población y sólo si ésta es homogénea, la representatividad de tal muestra puede ser satisfactoria.(Azorín, 1972, pag. 4).

Muestreo probabilístico : Cuando puede definirse claramente el procedimiento para obtener, de la unidades de muestreo de la población, el conjunto de muestras y calcularse la probabilidad de obtener cada una de ellas; se establecen un estimadores de los parámetros para cualquier muestra, y éstos resultan ser variables aleatorias. Por ejemplo, el promedio de los datos de la muestra.

Los diferentes diseños de muestreo probabilístico que se usan comúnmente son :

Muestreo aleatorio simple sin reemplazo

Muestreo aleatorio simple con reemplazo

Muestreo estratificado

Muestreo multietápico

Muestreo sistemático

Muestreo con probabilidad proporcional al tamaño y combinaciones de ellos.

Los procedimientos de estimación son algoritmos o fórmulas para estimar las características de la población con los datos provenientes de la muestra.

Para decidir el diseño de muestreo que será aplicado se toma en cuenta el costo (si los datos se obtienen de una pequeña fracción del total, los gastos son menores) que se relaciona con la precisión deseada, el trabajo de campo (el entrenamiento que deben recibir los encuestadores con respecto al objetivo de la encuesta, los métodos de medición aplicados, además de una supervisión de su trabajo), y las restricciones con respecto al tiempo

La decisión de escoger un diseño muestral está a cargo de un equipo formado por expertos en materia estadística, y especialistas en el campo donde se extrae la muestra; por ejemplo, en un estudio de tipo social, lo más común sería sociólogos.

Será de mucho valor, obtener información disponible sobre experiencias adquiridas en otras encuestas, detalles de logística, geografía, comunicación y perfil sociológico de la población a encuestar.

Una vez acordado el diseño, es importante asegurarse que los datos sean recabados de acuerdo al tipo de muestreo. Repasar el mecanismo de integración de las observaciones en el que están consideradas la captura, codificación, validación y transformación para obtener en forma oportuna y consistente los reportes y análisis estadísticos.

A continuación se describen los diseños de muestreo que serán utilizados en el presente trabajo: el muestreo aleatorio simple, el muestreo por conglomerados, el muestreo sistemático, y el muestreo estratificado.

1.2 Muestreo aleatorio simple (MAS)

Es el más sencillo de todos los diseños muestrales, consiste en etiquetar a los N elementos de la población y escoger al azar n para formar una muestra, cada una de las muestras tiene la misma probabilidad de ser escogida. Se obtiene de la característica bajo estudio un promedio, proporción o valor que estima al de la población.

Consideremos a la población $\mathcal{P}$ como el conjunto

$$\mathcal{P} = \{s_1, s_2, \ldots, s_N\}$$

donde cada s_i que pertenece a $\mathcal{P}$, representa al sujeto (persona, animal, objeto) o conjunto de ellos del que se estudia una característica o atributo

$$\mathcal{P}' = \{ Y_1, Y_2, \ldots, Y_i, \ldots, Y_N \}$$

es el conjunto de valores de la población donde puede establecerse una convención por la que cada $Y_i \in \mathbf{R}$ con $i = 1, 2, \ldots, N$.

La selección de los n sujetos en la muestra puede obtenerse al utilizar una tabla de números aleatorios o algún mecanismo digital como una calculadora. Algunos cuestionamientos surgen de la manera de escoger la muestra: ¿ cuál debe ser su tamaño?, los elementos de la muestra ¿ deben o no devolverse a la población antes de elegir el siguiente?, ¿qué tan costoso o difícil será obtener la muestra? y ¿qué tan preciso es el método para garantizar un valor aproximado al de la población ?.

A continuación se describen las dos maneras más comunes de escoger un muestra aleatoria.

1.2.1 Muestreo aleatorio simple sin reemplazo

Es el diseño muestral donde cada unidad de la población tiene la misma probabilidad de escogerse la primera vez. Después de seleccionarla y medir su característica no se regresa a la población y otra unidad se escoge entre las restantes, el proceso continúa hasta completar la muestra.

Una muestra de tamaño n es una serie de selecciones consecutivas de elementos de la población cuyos valores observados son $y_1, y_2, \ldots, y_i, \ldots, y_n$, la probabilidad de escoger una unidad en la primera selección es igual a $\frac{1}{N}$. Al no reemplazar la unidad s, en la población la probabilidad de escoger una unidad de las N-1 restantes es igual a $\frac{1}{N-1}$.

Una propiedad importante en el muestreo aleatorio simple sin reemplazo (MASSR) es que la probabilidad de escoger una unidad específica de la población en cada elección es igual a la probabilidad de escogerla en la primera selección.

Se denotará una muestra aleatoria simple sin reemplazo como m.a.s.s.r.

1.2.2 Muestreo aleatorio simple con reemplazo

Es el diseño muestral donde cada unidad seleccionada se introduce nuevamente en la población, de manera que pueda seleccionarse de nuevo. El proceso se repite hasta completar la n unidades de la muestra.

La población y los valores poblacionales asociados al muestreo aleatorio simple con reemplazo (MASCR) son los mismos que en el MASSR, pero la probabilidad de escoger una unidad en cualquier selección es igual a $\frac{1}{N}$.

Se utilizará la misma notación que toma Sukhatme (1984); los valores de la población se denotan con letras mayúsculas, y los de la muestra con minúsculas con el mismo subíndice i.

Cuando se tome una muestra en un MASCR se le denotará por m.a.s.c.r.

1.2.3 Parámetros poblacionales en un muestreo aleatorio simple

Las propiedades o atributos que se desean medir en una población se llaman *parámetros poblacionales*, tales como: la preferencia electoral, promedio de edad en un grupo de personas, el número total de hectáreas sembradas en una región, la proporción de ingreso / gasto de los habitantes del D.F., etc.

Como los parámetros son funciones de las características de la población, los denotaremos por $\theta = \theta(Y_1, Y_2, \ldots, Y_i, \ldots, Y_N)$ donde θ puede ser lineal o no, Y_i es la característica medida en el elemento *i-ésimo* de la población.

A continuación se presenta una lista de los parámetros más utilizados.

PARÁMETROS POBLACIONALES

Parámetro poblacional	Descripción	
$\bar{Y} = \dfrac{\sum_{i=1}^{N} Y_i}{N}$	media poblacional	(1.2.3.a)
$T_Y = \sum_{i=1}^{N} Y_i$	total poblacional	(1.2.3.b)
$R = \dfrac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i}$	razón poblacional de Y sobre X	(1.2.3.c)

$$S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1} \qquad \text{suma de cuadrados medios} \qquad (1.2.3.d)$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N} \qquad \text{varianza poblacional} \qquad (1.2.3.e)$$

1.2.4 Estimadores de los parámetros poblacionales en un muestreo aleatorio simple

Cuando es de interés conocer el valor de un parámetro de la población se realiza una estimación a partir de los valores que se obtienen de la muestra. Un estimador es una función de lo valores y_i, denotaremos a $\hat{\theta} = \hat{\theta}(y_1, y_2, \ldots y_i, \ldots, y_n)$ como un estimador de θ. Enseguida se presenta una lista de los estimadores más usuales.

Estimador	Descripción	
$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$	media muestral	(1.2.4.a)
$t_y = N\bar{y}$	total muestral	(1.2.4.b)
$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$	suma de cuadrados medios	(1.2.4.c)
$r_Y = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$	razón muestral de Y sobre X	(1.2.4.d)

Si un estimador cumple que su valor esperado sea el parámetro poblacional, se dice que es insesgado. Formalmente tenemos:

Def. Sea θ parámetro de la población, si $E(\hat{\theta}) = \theta$ entonces $\hat{\theta}$ es un *estimador insesgado* de θ. Se pueden deducir los siguientes resultados:

a) $\bar{y}$ es un estimador insesgado de $\bar{Y}$

b) s^2 es un estimador insesgado de S^2

c) N$\bar{y}$ es un estimador insesgado de T_Y.

Cuando no se tiene un estimador insesgado no necesariamente es inútil, sobre todo cuando se trabaja con muestras grandes.

Un estimador $\hat{\theta}$ es un *estimador consistente* de θ si el estimador es igual al parámetro cuando la muestra toma toda la población (Sukhatme, 1984, pag. 14), o cuando el tamaño de la muestra.tiende a infinito

La varianza V de un estimador es la varianza de la población de valores $\hat{\theta}$ que resultarían si se repitiese el muestreo muchas veces, mide su precisión alrededor del valor promedio poblacional. Si es pequeña, la probabilidad de obtener una gran diferencia con respecto al valor poblacional es también reducida.

Cuando los valores de la población son muy parecidos u homogéneos la varianza es menor que en una con gran diversidad en sus características.

Los siguientes resultados son válidos en un MAS.

$$V(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{S^2}{n} \quad \text{en un muestreo simple sin reemplazo} \qquad (1.2.4.e)$$

$$V(\bar{y}) = \left(\frac{N-1}{N}\right)\frac{S^2}{n} \quad \text{en un muestreo simple con reemplazo} \qquad (1.2.4.f)$$

En ambos diseños, la varianza depende del valor de S^2 que puede estimarse por s^2. Se observa que la varianza del estimador de la media en un MASSR es menor que la de un MASCR.

La precisión de los estimadores en el MAS depende de la diversidad de la caracteristica medida y está sujeta a la fluctuación en el muestreo. La varianza de $\bar{y}$ depende del tamaño de la muestra, de la varianza de la población, y del tamaño N. Aunque si n/N es pequeño, no.

Las fórmulas para los errores estándar (suma de cuadrados) del estimador de la media y total de la población se usan principalmente para comparar su precisión con respecto a otros diseños de muestreo y para estimar el tamaño de la muestra .

Los parámetros más frecuentemente estimados en un diseño muestral son la media poblacional y el total.

Generalmente si $\bar{y}$ es el promedio de las observaciones de una muestra simple aleatoria, se tiene que hay un pequeño riesgo α que se está dispuesto a correr de que el error real sea mayor que d; es decir, de que

$$\Pr(|\bar{y} - \bar{Y}| \geq d) = \alpha$$

en donde d es el margen de error elegido y α es una probabilidad pequeña. Si se supone que y se distribuye normalmente, se tiene de la fórmula (1.2.4.e)

$$\sqrt{V(\bar{y})} = \sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

y la fórmula que relaciona a n con el grado de precisión deseado es

$$d = t\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

donde t es la abcisa del punto que separa un área α en los extremos de la distribución. Al resolver para n.

$$n = \frac{\left(\frac{tS}{d}\right)^2}{1 + \frac{1}{N}\left(\frac{tS}{d} - \right)^2}$$

si N es grande, se puede tomar como primera aproximación a

$$n_0 = \left(\frac{tS}{d}\right)^2 = \frac{S^2}{V}$$

Si n_0/n es pequeña, se calcula a n como

$$n = \frac{n_0}{1 + \left(\frac{n_0}{n}\right)} \qquad (1.2.a)$$

Además de estimar la media o total poblacional, es de interés hacerlo con la proporción del número de unidades de muestreo cuya caracterísitca posee un atributo en particular. Por ejemplo, si se trata de conocer la proporción de votos a favor de un determinado partido, se puede definir una variable dicotómica para cada voto de la siguiente manera : 1 si es a favor del partido, 0 si no lo es.

Si P es la proporción de la característica en la población, el tamaño de muestra n se puede calcular como :

$$n = \frac{\frac{t^2 PQ}{d^2}}{1 + \frac{1}{N}\left(\frac{t^2 PQ}{d^2} - 1\right)} \qquad (1.2.b)$$

donde t es la abscisa de la curva normal que corta a un área de α en las colas de la distribución, N es el tamaño de la población y Q=1-P. (Cochran, 1977, pag. 75).

En la práctica se sustituye un estimador p de P .Si N es grande, una primera aproximación es $n_0 = \frac{t^2 pq}{d^2} = \frac{pq}{V}$, si n_0/n no es despreciable se utiliza a (1.2.a)

La decisión del tamaño de muestra es una de las partes importantes del diseño. Está determinada por la precisión deseada y el costo que implica el escoger un número específico, si la muestra es muy grande podría ser más costosa sin ganar mucha precisión o perder eficiencia por reducirse.

En las siguientes secciones se analizarán otros diseños muestrales.

1.3 Muestreo de conglomerados

Un muestreo por conglomerados consiste en agrupar a los elementos de una población en conjuntos llamados conglomerados (cada conglomerado es una unidad de muestreo), tomar una muestra de conglomerados de acuerdo a un diseño de muestreo y censar en los elementos del conglomerado la característica bajo estudio para obtener una estimación del valor poblacional.

Las unidades de muestreo más pequeñas en que es posible dividir una población se llaman unidades elementales. Pueden reunirse en diferentes grupos para formar unidades denominadas conglomerados. Por ejemplo, si el muestreo a realizar considera a la población de toda la República Mexicana, las unidades más pequeñas o elementales son los habitantes, algunos de los conglomerados que pueden considerarse son : familias, cuadras, secciones electorales, distritos electorales, colonias, municipios o estados.

Un diseño de muestreo de conglomerados puede utilizarse cuando no se dispone de una lista confiable de unidades elementales para considerarlas unidades de muestreo, por lo que se agrupan. Cada unidad elemental debe pertenecer a uno sólo de los conglomerados sin omisiones o duplicaciones, con la finalidad de evitar sesgos y, además, la unión de todos los conglomerados es el conjunto de todas las unidades de muestreo de la población, lo que evita que ocurran errores de omisión y duplicación.

Al estudiar el comportamiento electoral es difícil conocer la preferencia de un elector al sufragar, pero es posible conocer la información de todos los que votaron al cierre de la casilla.

Los conglomerados escogidos en la muestra se enumeran completamente, una de las ventajas es que las listas de conglomerados son más fáciles de conseguir, lo que reduce costos.

En una ciudad es más fácil obtener una lista de viviendas que una lista de sus moradores, además, una muestra aleatoria de viviendas se dispersa mucho más en una ciudad grande (lo que la hace más costosa) que una muestra equivalente por cuadras o barrios, para cada conglomerado se tendrían varias viviendas juntas. El costo de la muestra aleatoria sería mayor, aunque cubriría más uniformemente la ciudad y se podría obtener una mayor precisión.

Si se consideran el costo y la precisión, el muestreo por conglomerados obtiene, en general, mejores resultados. Una pregunta que surge es cómo deben ser formados los conglomerados para obtener una mejor precisión.

Para ganar precisión al formar los conglomerados debe procurarse que tengan una gran heterogeneidad interna, pero que guarden similitud entre ellos como se verá más adelante. Por lo general los conglomerados se forman al unir familias, viviendas, etc. que están geográficamente juntas con lo que se obtiene homogeneidad interna al censar sus características, por lo que se pierde precisión.

La eficiencia del diseño depende (Sukhatme, 1984, pag. 270), entre otros aspectos, del tamaño del conglomerado y de la forma de establecer la unidad de muestreo.

En la siguiente sección se presenta el caso en que los conglomerados tienen igual tamaño. Se describen los estimadores de la media y varianza .

1.3.1 Conglomerados de igual tamaño

El caso en que los conglomerados son de igual tamaño se presenta primero, aunque en la práctica cuando se realiza un muestreo es difícil conservar el mismo tamaño.

Para este diseño de muestreo usaremos la notación siguiente (Sukhatme, 1984, pag. 271). N es el número de conglomerados en la población y n el tamaño de una m.a.s.s.r. de conglomerados.

Y_{ij} es el valor de la característica bajo estudio para el *j-ésimo* elemento en el *i-ésimo* conglomerado de la población, $i=1,2,...,N$; $j=1,2,...,M$.

M es el tamaño del conglomerado, número de elementos en cada uno

$\overline{Y}_{i.} = \dfrac{\sum_{j=1}^{M} Y_{ij}}{M}$ la media por elemento del *i-ésimo* conglomerado $i=1,2,...,n$ (1.3.1.a)

$\overline{\overline{Y}} = \dfrac{\sum_{i=1}^{N} \overline{Y}_{i.}}{N}$ la media de las medias de los conglomerados (1.3.1.b)

$\overline{Y} = \dfrac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij}$ la media poblacional (1.3.1.c)

$S_i^2 = \dfrac{1}{M-1}\sum_{j=1}^{M}(Y_{ij} - \overline{Y}_{i.})^2$ el cuadrado medio entre elementos dentro del *i-ésimo* conglomerado $i=1,2,...,n$ (1.3.1.d)

$S_w^2 = \frac{1}{N}\sum_{i=1}^{N} S_i^2$ el cuadrado medio dentro de los conglomerados (1.3.1.e)

$S_b^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_i.-\overline{\overline{Y}})^2$ el cuadrado medio entre medias de conglomerados (1.3.1.f)

1.3.1.1 Estimador de la media , varianza del estimador y estimador de su varianza

Cuando el tamaño de los conglomerados es igual $\overline{\overline{y}} = \frac{\sum_{i=1}^{n}\overline{y}_i.}{n}$ la *media de las medias*, es un estimador insesgado de $\overline{\overline{Y}} = \overline{Y}$ y su varianza es :

$$V(\overline{\overline{y}}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2 \qquad (1.3.1.1.a)$$

y como $E(s_b^2) = S_b^2$ se tiene que

$$\hat{V}(\overline{\overline{y}}) = \left(\frac{1}{n} - \frac{1}{N}\right)s_b^2 \qquad (1.3.1.1.b)$$

es un estimador de su varianza.

La varianza de los estimadores tiende a aumentar al ser los conglomerados generalmente homogéneos internamente, su distribución física en la población es habitualmente geográfica y no aleatoria (Kish, 1965, pag. 197). La manera de procurar revertir esta tendencia se presenta a continuación.

Para conseguir una mayor eficiencia en un muestreo de conglomerados con igual tamaño, los conglomerados deben agruparse de manera que la variación entre las medias de las características de cada conglomerado sea lo más pequeña posible (los conglomerados deben ser parecidos), mientras que la variación interna debe ser tan grande como se pueda (los conglomerados deben ser heterogéneos internamente), resultado que surge del siguiente análisis .

La eficiencia relativa (E.R.) de la varianza de la media del muestreo por conglomerados con respecto al MAS es :

$$E.R.(\bar{\bar{y}})=\frac{V_{mas}(\bar{y})}{V_{cong}(\bar{\bar{y}})}=\frac{(\frac{1}{nM}-\frac{1}{NM})S^2}{(\frac{1}{n}-\frac{1}{N})S_b^{\,2}}=\frac{S^2}{MS_b^{\,2}} \qquad (1.3.1.1.c)$$

donde N es el tamaño de la población, n el de la muestra, M el del conglomerado, $S_b^{\,2}$ la suma de cuadrados medios entre los conglomerados y $\bar{\bar{y}}$ la media de las medias.

La eficiencia relativa es grande si el denominador $MS_b^{\,2}$ disminuye. La eficiencia del muestreo por conglomerados se incrementa en la medida en que la suma de los cuadrados medios entre los conglomerados decrece.

$MS_b^{\,2}=\frac{1}{N-1}\left\{(NM-1)S^2-N(M-1)S_w^{\,2}\right\}$ y para que $MS_b^{\,2}$ decrezca $S_w^{\,2}$ debe incrementarse.

Los conglomerados deben ser heterogéneos internamente y homogéneos entre ellos para aumentar la E.R. con respecto al MAS.

1.3.2 Conglomerados de diferente tamaño

El caso se presenta cuando los conglomerados son de diferente tamaño, en la práctica es lo que con más frecuencia ocurre. Para este tipo de muestreo se tiene la siguiente notación.

N es el número de conglomerados

M_i es el número de elementos del conglomerado $i=1,2,\ldots,N$

$$\overline{M}=\frac{\sum_{i=1}^{N}M_i}{N}$$ tamaño promedio de un conglomerado (1.3.2.a)

$$\overline{Y}_{i\cdot}=\frac{\sum_{j=1}^{M_i}Y_{ij}}{M_i}$$ media por elemento del *i-ésimo* conglomerado $i=1,2,\ldots,n$ (1.3.2.b)

$$\overline{\overline{Y}}=\frac{\sum_{i=1}^{N}\overline{Y}_{i\cdot}}{N}$$ media de las medias de los conglomerados (1.3.2.c)

$$U_{ij}=\frac{M_iY_{ij}}{\overline{M}} \qquad (1.3.2.d)$$

$$\overline{U}_{i\cdot}=\frac{\sum_{j=1}^{M_i}U_{ij}}{M_i}=\frac{M_i\overline{Y}_{i\cdot}}{\overline{M}} \qquad (1.3.2.e)$$

$$V_i=\frac{M_i}{\overline{M}} \qquad (1.3.2.f)$$

$$\overline{Y}=\frac{\sum_{i=1}^{N}M_i\overline{Y}_{i\cdot}}{\sum_{i=1}^{N}M_i}=\frac{\sum_{i=1}^{N}M_i\overline{Y}_{i\cdot}}{N\overline{M}}=\frac{\sum_{i=1}^{N}\overline{U}_i}{N}=\overline{\overline{U}} \qquad (1.3.2.g)$$

$$\overline{V}=\frac{\sum_{i=1}^{N}V_i}{N} \qquad (1.3.2.h)$$

1.3.2.1 Tres estimadores de la media, varianzas de los estimadores y estimadores de sus varianzas.

Existen varios estimadores para estimar a la media cuando los conglomerados tienen tamaño diferente, en un MASSR existen al menos cuatro estimadores (Sukhatme, 1984, pag. 291) que son los siguientes.

$$a)\ \hat{\bar{Y}} = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n} \quad \textit{media de medias,} \qquad (1.3.2.1.a)$$

es un estimador sesgado

$$b)\ \hat{\bar{Y}}_u = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}} = \bar{\bar{u}} \quad \textit{estimador insesgado} \text{ de la media} \qquad (1.3.2.1.b)$$

$$c)\ \hat{\bar{Y}}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i} = \frac{\bar{\bar{u}}}{\bar{v}} \quad \textit{estimador de razón} \qquad (1.3.2.1.c)$$

es sesgado pero consistente y no requiere conocer $\bar{M}$

$$d)\ \hat{\bar{Y}}_l = \bar{\bar{u}} + \hat{\beta}_{uv}(1-\bar{v}) \text{ con } \hat{\beta}_{uv} = S_{uv} / S_b^2 \quad \textit{estimador de regresión} \qquad (1.3.2.1.d)$$

es sesgado , aunque el sesgo es pequeño si la relación entre u y v es lineal

$$\text{donde } \bar{y}_{i\cdot} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i} \text{ con } i=1,2..,n \qquad (1.3.2.1.e)$$

Observación : Si en *b*) se estima a $\bar{M}$ con $\bar{M}_n = \frac{\sum_{i=1}^{n} M_i}{n}$, el estimador que se obtiene es el estimador de razón.

El estimador media de medias es un estimador sesgado y su error cuadrático medio es :

$$a')\qquad ECM(\hat{\bar{Y}}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^2 \qquad (1.3.2.1.f)$$

$$\text{donde } S_b^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (\bar{Y}_{i\cdot} - \bar{\bar{Y}})^2 \qquad (1.3.2.1.g)$$

un estimador insesgado del error cuadrático medio de este estimador es :

$$\widehat{ECM}(\hat{\bar{Y}}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_b^2 \qquad (1.3.2.1.h)$$

$$\text{donde } s_b^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(\bar{y}_{i\cdot}-\bar{\bar{y}})^2 \qquad (1.3.2.1.i)$$

La varianza del estimador insesgado es

b')
$$V(\hat{\bar{Y}}_u) = \left(\frac{1}{n}-\frac{1}{N}\right)S_u^2 \qquad (1.3.2.1.j)$$

$$\text{donde } S_u^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(\frac{M_i\bar{Y}_{i\cdot}}{\bar{M}}-\bar{\bar{Y}}\right)^2 \qquad (1.3.2.1.k)$$

y un estimador insesgado de la varianza es :

$$\hat{V}(\hat{\bar{Y}}_u) = \left(\frac{1}{n}-\frac{1}{N}\right)s_u^2 \qquad (1.3.2.1.l)$$

$$\text{donde } s_u^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(\bar{u}_{i\cdot}-\bar{\bar{u}})^2 \qquad (1.3.2.1.m)$$

El error cuadrático medio del estimador de razón es aproximadamente

c')
$$ECM(\hat{\bar{Y}}_r) \doteq \left(\frac{1}{n}-\frac{1}{N}\right)S_{u,v}^2 \qquad (1.3.2.1.n)$$

$$\text{donde } S_{u,v}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{U}_{i\cdot}-\bar{Y}V_i)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\frac{M_i^2}{\bar{M}_2}(\bar{Y}_{i\cdot}-\bar{Y})^2 \qquad (1.3.2.1.ñ)$$

1.4 Muestreo sistemático

En un muestreo sistemático se supone que las unidades de la población se numeran del 1 al N, se considera que $\frac{N}{n}=k$ para algún entero k (a k se le llama intervalo de muestreo), N es el tamaño de la población y n el de la muestra. Para seleccionar una muestra de n unidades se toma al azar un número i (arranque aleatorio) entre 1 y k, la muestra contiene n valores de las unidades con los números seriados :

$$i,\ i+k,\ i+2k,\dots,\ i+(n-1)k$$

la selección del primer elemento determina automáticamente toda la muestra; esto es equivalente a dividir a los N elementos en k clases C_1 , $C_2,\dots,C_k$ donde C_i consiste de y_i, y_{i+k},

..., $y_{i+(n-1)k}$ y se elige al azar alguno de los C_i, por lo tanto, es equivalente al muestreo de conglomerados de igual tamaño como se describe a continuación.

Una manera de considerar el muestreo sistemático como un muestreo de conglomerados de igual tamaño en donde se toma una muestra de tamaño uno es la siguiente.

En un muestreo sistemático $k=N/n$, la población $\mathcal{P} = \{s_1, s_2, \ldots, s_N\}$ puede reunirse en k conglomerados de tamaño n de la siguiente manera:

$C_1=\{s_1, s_{1+k}, s_{1+2k},\ldots,s_{1+(n-1)k}\}$

$C_2=\{s_2, s_{2+k}, s_{2+2k},\ldots,s_{2+(n-1)k}\}$

.

.

.

$C_k=\{s_k, s_{2k}, s_{3k},\ldots,s_{nk}\}$

Se elige al azar un conglomerado C_i, $C_i=\{s_i, s_{i+k}, s_{i+2k},\ldots,s_{i+(n-1)k}\}$ y al censarlo se obtienen los valores $y_i, y_{i+k}, y_{i+2k},\ldots, y_{i+(n-1)k}$ lo que es equivalente a tomar una muestra de tamaño n en un muestreo sistemático. En consecuencia, la varianza del estimador de la media en un muestreo sistemático es la misma que para una muestra de un sólo conglomerado (una sóla observación), por lo que no es posible calcularla bajo este enfoque. Generalmente se utiliza la varianza de los estimadores de un MAS.

Una muestra bajo un diseño de muestreo sistemático es más fácil de obtener, puede ejecutarse prácticamente sin equivocarse y en menos tiempo. Una de las desventajas del diseño ocurre cuando se tiene una formación periódica de las características, y el intervalo k de muestra escogido coincide con el periodo de los datos.

Algunos estudios sobre muestreo sistemático se han realizado para medir la conveniencia de aplicarlo en forma alternativa al MAS (Madow, 1946), en este artículo se aborda su relación con el MAS, el muestreo estratificado (que se define más adelante) y el de conglomerados. Las relaciones se mencionan a continuación.

La eficiencia del diseño sistemático comparada con el MAS es errática, ya que no tiene un comportamiento determinado. El muestreo sistemático es más eficiente que el muestreo estratificado para algunos tamaños de muestra, pero es menos eficiente en algunos otros, depende del orden de los valores Y_i. en el marco de muestra.

Para los datos del artículo (Madow,1946) el muestreo estratificado es más eficiente que el MAS para todos los valores de n, pero no alcanza el grado de eficiencia que un muestreo sistemático. Cuando el muestreo sistemático es mejor que el muestreo estratificado su eficiencia es 2 veces más grande, mientras que en la mayoria de los casos en que el muestreo sistemático es menos eficiente que el muestreo estratificado, lo es por una pequeña diferencia.

Madow reporta algunos ejemplos numéricos en los que el muestreo estratificado puede tener mayor ventaja. Concluye que si ρ (el coeficiente de correlación intraclase de los elementos del conglomerado) es negativo, el muestreo de conglomerados tiene una varianza menor que el MAS y si ρ es positivo sucede lo contrario, pero cuando es cercano a cero los dos diseños tienen aproximadamente la misma varianza, aunque la del muestreo de conglomerados es la mayor.

En el artículo mencionado se comparan resultados numéricos de la eficiencia del MAS con el muestreo sistemático y el estratificado para diferentes tamaños de muestra. Para los datos presentados en el artículo en particular, la eficiencia del muestreo estratificado es mayor la del MAS cuando se incrementa el tamaño de muestra, pero no alcanza la eficiencia que en algunos casos consigue el muestreo sistemático.

Se describe en la siguiente sección un caso particular del muestreo sistemático.

1.4.1 Muestreo sistemático centrado, un estimador de la media

En un muestreo sistemático es posible formar k valores de la media $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$, donde $\bar{y}_i$ es la media de la muestra cuyo primer elemento es y_i.

El muestreo sistemático centrado considera dos casos para definir el estimador a partir del valor de k; si k es impar, el ***estimador de muestreo sistemático centrado*** $\bar{y}_c$ es $\bar{y}_{\frac{(k+1)}{2}}$, y si k es par se define como $\bar{y}_c = \bar{y}_{\frac{k}{2}}$.

Uno de los artículos que compara el muestreo sistemático centrado y el muestreo sistemático con arranque aleatorio (Madow, 1952) reporta que cuando las características de las poblaciones son monótonas decrecientes, esto es $\bar{y}_1 \leq \bar{y}_2 \leq \ldots \bar{y}_k$, el muestreo sistemático centrado es más eficiente que el muestreo sistemático con arranque aleatorio; también señala que si las características en una población son monótonas (es decir, $\bar{y}_1 \leq \bar{y}_2 \leq \ldots \bar{y}_k$ ó $\bar{y}_1 \geq \bar{y}_2 \geq \ldots \bar{y}_k$), el muestreo sistemático centrado es más eficiente que el muestreo sistemático con arranque aleatorio.

En ese artículo, Madow cita la existencia de casos en los que un muestreo estratificado es más eficiente que el muestreo sistemático con arranque aleatorio, menciona las condiciones (llamadas de concavidad ascendente y correlación decreciente) bajo las que Cochran (1946, pags. 164-177) probó que ocurre lo contrario.

1.5 Muestreo estratificado

En un MAS la varianza de los estimadores depende del tamaño de muestra y la variabilidad de las características, sus estimaciones pueden diferir debido a la desigualdad de los valores de la muestra.

Las poblaciones en las que se aplica una encuesta son generalmente heterogéneas con respecto a la característica de interés, la varianza del estimador de la media o total es habitualmente grande y la precisión que se consigue es poca.

Uno de los métodos para lograr que se reduzca la variabilidad al estimar un parámetro es dividir a la población en subconjuntos que sean similares internamente a partir de un criterio establecido (nivel económico, vecindad geográfica, etc.). Los grupos se llaman estratos, la división se aplica para que las características sean también muy similares dentro de cada estrato y obtener una varianza reducida. Además se busca que los estratos sean lo más diferentes entre si. Una vez que se tienen establecidos los estratos, se escoge una muestra de cada uno de manera independiente. A este método se le conoce como muestreo estratificado, si se escoge de cada estrato un muestra aleatoria simple el procedimiento es llamado muestreo estratificado aleatorio.

Existen varias justificaciones para estratificar una población, algunas son: la conveniencia administrativa de poder encuestar en los estratos con un mejor control y reducir así los costos; los elementos de la población pueden tener una característica que haga conveniente considerar estratos, por ejemplo, para estudiar la atención médica, a los hospitales se les puede clasificar de acuerdo a su tamaño en varios estratos. Se considera que el criterio logra homogeneizar la característica bajo estudio al agruparlos así. La atención médica de los hospitales va a ser muy variable entre estratos, aunque se espera que lo sea menos dentro de los estratos.

Una población de sujetos es posible dividirla bajo un criterio que clasifique a sus elementos en grupos parecidos, esto permite reducir la variabilidad de sus respuestas y se consigue mejorar la precisión de sus estimadores. Esto es, la estratificación puede dar lugar a una ganancia en precisión en los estimadores de la característica de interés en toda la población.

El diseño de muestreo estratificado es recomendable para obtener estimaciones de una población grande como un país o estado. La división política por estados o municipios permite clasificarlos en estratos conocidos

El muestreo estratificado es un procedimiento común en investigaciones por muestreo, garantiza una representación adecuada de todos los estratos. La selección de unidades, la distribución y supervisión del trabajo se simplifican. En el MAS no puede asegurarse una representación de esa índole, la muestra puede estar formada por elementos que representen más a algunos estratos que a otros.

Otro método de construir los estratos es usar la información sobre algún factor correlacionado con la característica de interés como la base para la estratificación. Por ejemplo, la producción de una fábrica deberá estar correlacionada con su tamaño, la estratificación con este criterio conseguirá una mejor precisión de los estimadores de la producción.

Cuando existen subdivisiones que se consideran importantes para realizar estimaciones separadas con una precisión determinada, se les trata como una población por sí mismas y se les llama *dominios de estudio*. Un ejemplo de ello son las grandes ciudades del país, que pueden dividirse en dos clases: una, con habitantes propietarios de casa; otra, con habitantes que la rentan, y se desea obtener información que sirve como indicador del rechazo o aceptación de un determinado producto o de algún partido político.

En el muestreo estratificado la población $\mathcal{P}$ de tamaño N se divide en L regiones o estratos

$$\mathcal{P}=\left\{\left\{s_{11},s_{12},\ldots,s_{1N_1}\right\},\left\{s_{21},s_{22},\ldots,s_{2N_2}\right\},\ldots,\left\{s_{L1},s_{L2},\ldots,s_{LN_L}\right\}\right\}$$

donde s_{ij} representa al sujeto j-ésimo del i-ésimo estrato $(i=1,2,..,L\ ;\ j=1,2,...,N_i)$

Cada estrato incluye sujetos que guardan similitud de acuerdo al criterio establecido, las características de interés de la población son :

$$\mathcal{P}'=\left\{\left\{Y_{11},Y_{12},\ldots,Y_{1N_1}\right\},\left\{Y_{21},Y_{22},\ldots,Y_{2N_2}\right\},\ldots,\left\{Y_{L1},Y_{L2},\ldots,Y_{LN_L}\right\}\right\}$$

donde Y_{ij} es la característica bajo estudio medida en la j-ésima *unidad* de muestreo del i-ésimo estrato , $(i=1,2,...,L\ ;\ j=1,2,..,N_i)$

Se usará la notación siguiente:

N_i es el número de unidades en el *i-ésimo* estrato i=1,2,...,L

$W_i= N_i/N$ el peso de cada estrato con respecto a la población i=1,2,...,L

$$\overline{Y}_i = \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}$$ la media basada en las N_i características de cada estrato.Media de cada estrato i=1,2,...,L (1.5.b)

$$S_i^2 = \frac{\sum_{j=1}^{N_i}(Y_{ij} - \overline{Y}_i)^2}{N_i - 1}$$ suma de cuadrados medios basado en los N_i valores de la característica del i-ésimo estrato i=1,2,..,n (1.5.c)

$$\overline{Y} = \frac{\sum_{i=1}^{L}\sum_{j=1}^{N_i} Y_{ij}}{N} = \frac{\sum_{i=1}^{L} N_i\overline{Y}_i}{N} = \sum_{i=1}^{L} W_i\overline{Y}_i$$ la media poblacional (1.5.d)

Una muestra $\mathcal{M}_i$ de tamaño n_i se selecciona en forma independiente (mediante un diseño de muestreo como el MASSR, MASCR, muestreo sistemático, etc.) en cada estrato. Se calcula su media y se pondera para formar una estimador de la media poblacional, análogamente se procede para estimar las varianzas de las medias, ponderándolas para conseguir un estimador de la varianza. Cuando se escoge la muestra en cada estrato mediante un MAS,

se denota a :

n_i	el tamaño de muestra en el *i-ésimo* estrato	$i=1,2,...,L$
$f_i= n_i /N_i$	la fracción de muestreo en el *i-ésimo* estrato.	$i=1,2,...,L$

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$ la media muestral del *i-ésimo* estrato. $i=1,2,...,L$ (1.5.e)

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1}$$ suma de cuadrados medios del *i-ésimo* estrato basada en n_i valores de la muestra. $i=1,2,...L$ (1.5.f)

Si en cada estrato se escoge una m.a.s.s.r. en forma independiente, se generan los siguientes resultados para la media de cada estrato :

$E(\bar{y}_i) = \bar{Y}_i$ la media muestral de cada estrato es un estimador insesgado para cada estrato $i=1,2,...,L$

$V(\bar{y}_i) = \frac{(N_i - n_i)}{N_i}\frac{S_i^2}{n_i}$ la varianza de la media muestral en cada estrato

$E(s_i^2) = S_i^2$ la suma de cuadrados medios en la muestra en cada estrato es un estimador insesgado de la suma de cuadrados en cada estrato.

se obtiene un estimador insesgado de la media poblacional mediante

$$\bar{y}_{st} = \sum_{i=1}^{L} W_i \bar{y}_i \qquad (1.5.g)$$

que tiene como varianza a :

$$V(\bar{y}_{st}) = \sum_{i=1}^{L} W_i^2 \frac{(N_i - n_i)}{N_i}\frac{S_i^2}{n_i} \qquad (1.5.h)$$

y su estimador insesgado correspondiente es :

$$\hat{V}(\bar{y}_{st}) = \sum_{i=1}^{L} W_i^2 \frac{(N_i - n_i)}{N_i} \frac{s_i^2}{n_i} \tag{1.5.i}$$

Las condiciones en que el muestreo estratificado es mejor que el MAS se analizan más adelante.

El tamaño óptimo de muestra está en función de su costo o su precisión. Se utilizará un costo fijo para obtener el tamaño óptimo de muestra al minimizar la varianza.

Si consideramos a c_i como el costo por muestreo en el *i-ésimo* estrato, C el costo total de la encuesta y $W_i = N_i / N$ el peso de los estratos, el valor mínimo de la $V(\bar{y}_{st})$ para un costo C_o fijo ocurre cuando

$$n_i = \frac{C_0 W_i S_i}{\sqrt{c_i} \sum_{i=1}^{L} W_i S_i \sqrt{c_i}} \quad \text{llamada } \textit{asignación óptima} \tag{1.5.j}$$

Cuando c_i es igual para todos los estratos y el costo de la encuesta es proporcional al tamaño de la muestra, la asignación óptima se reduce a :

$$n_i = n \frac{W_i S_i}{\sum_{i=1}^{L} W_i S_i} \quad \text{llamada } \textit{asignación de Neyman} \tag{1.5.k}$$

Si n_i es proporcional a W_i se le denomina *asignación proporcional* y

$$n_i = nW_i$$

Para comparar el muestreo estratificado bajo las asignaciones descritas con el MAS se establece primeramente la equivalencia de la asignación proporcional:

Al sustituir $n_i = nW_i$ en (1.5.h), se tiene

$$V(\bar{y}_{st})_P = \sum_{i=1}^{L} W_i^2 \frac{N_i - n_i}{N_i} \frac{S_i^2}{n_i} = \frac{N-n}{Nn} \sum_{i=1}^{L} W_i S_i^2 \tag{1.5.l}$$

donde el índice P significa la varianza bajo asignación proporcional.

En un MAS $V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S^2$, para tener a S^2 en función de las S_i^2 se requiere el siguiente resultado.

La suma de cuadrados total en la población puede a su vez expresarse utilizando la identidad que la divide en dos partes : a) suma de cuadrados dentro de los estratos y b) suma de cuadrados entre los estratos

$$\sum_{i=1}^{L}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y})^2 = \sum_{i=1}^{L}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i + \bar{Y}_i - \bar{Y})^2 = \sum_{i=1}^{L}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^{L} N_i(\bar{Y}_i - \bar{Y})^2$$

que puede escribirse como

$$(N-1)S^2 = \sum_{i=1}^{L}(N_i - 1)S_i^2 + \sum_{i=1}^{L} N_i(\bar{Y}_i - \bar{Y})^2 \qquad (1.5.m)$$

Suponiendo que N_i es lo suficientemente grande para permitir las aproximaciones $\frac{N_i - 1}{N_i} \cong 1$ y $\frac{N-1}{N} \cong 1$ se llega que

$$S^2 \cong \sum_{i=1}^{L} W_i S_i^2 + \sum_{i=1}^{L} W_i(\bar{Y}_i - \bar{Y})^2$$

al despejar y sustituir en (1.5.*l*) para que esté en términos de S^2, y restando la varianza de la media en un MASSR se obtiene que

$$V(\bar{y}) - V(\bar{y}_{st})_P = \frac{N-n}{Nn}\sum_{i=1}^{L} W_i(\bar{Y}_i - \bar{Y})^2 . \qquad (1.5.n)$$

esto significa que mientras más difieran los estratos en sus medias será mayor la ganancia en precisión del muestreo por asignación proporcional a la del MAS.

Para la asignación de Neyman, se obtiene la siguiente ecuación :

$$V(\bar{y}_{st})_P - V(\bar{y}_{st})_N = \frac{1}{n}\sum_{i=1}^{L} W_i(S_i - \bar{S}_w)^2 \qquad (1.5.ñ)$$

donde $\overline{S}_w = \sum_{i=1}^{L} W_i S_i$, que indica que mientras mayor diferencia exista entre las desviaciones estándar de los estratos, más grande será la ganancia en precisión de la asignación óptima sobre la asignación proporcional al tamaño.

Al comparar la varianza bajo asignación óptima en el muestreo estratificado, con la del MAS y sumar (1.5.n) con (1.5.p), se tiene que

$$V(\bar{y}) - V(\bar{y}_{st})_N = \frac{N-n}{Nn}\left\{\sum_{i=1}^{L} W_i(\overline{Y}_i - \overline{Y})^2 + \frac{N}{N-n}\sum_{i=1}^{L} W_i(S_i - \overline{S}_w)^2\right\} \qquad (1.5.o)$$

cuando esta cantidad es muy grande y positiva , el muestreo estratificado es mejor que el MAS.

El resultado muestra que para una estratificación eficiente, la población debe dividirse de manera que las diferencias entre las medias y desviaciones estándar de las características de los estratos sean tan grandes como sea posible; la mejor manera de hacerlo en la práctica es agrupar las unidades parecidas es decir, distribuyéndolas en estratos internamente homogéneos pero heterogéneos entre ellos.

Por último, si se compara la precisión del estimador de la media en un MAS con la de un muestreo estratificado que tiene una muestra repartida arbitrariamente entre los estratos

$$V(\bar{y}) - V(\bar{y}_{st}) = \sum_{i=1}^{l}\left(\frac{1}{n} - \frac{W_i}{n_i}\right) W_i S_i^2 + \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{L} W_i(\overline{Y}_i - \overline{Y})^2 \qquad (1.5.p)$$

el segundo término de la parte derecha de la ecuación es siempre positivo, pero el primero puede ser positivo, negativo o cero.

Es positivo cuando los n_i se escogen con asignación de Neyman, es cero si $n_i = nW_i$ o

$$n_i = n\frac{W_i S_i}{\sum_{i=1}^{L} W_i S_i}.$$

Cuando n_i se aleja de estas asignaciones, el primer término puede hacerse negativo y mayor que el segundo, con lo que el muestreo estratificado es *menos eficiente* que el MAS.

Estos resultados indican que debe considerarse en forma importante la asignación para cada muestra en los estratos.

El criterio de estratificación geográfica es un procedimiento habitualmente utilizado por suponer que los sujetos tienen problemática e intereses comunes; sin embargo, no necesariamente se tiene una certeza en la precisión, ya que las ganancias en precisión con esta estratificación son generalmente modestas (Cochran, 1977 , pag.102).

Las poblaciones humanas pueden estratificarse con base en el tamaño de su ciudad o localidad (al reunir en cada estrato una determinada cantidad de localidades con un número específico de habitantes), por sexo, factores socioeconómicos o electorales (distrito electoral).

La decisión de elegir el criterio de estratificación y el número de estratos es parte importante del diseño, ya que cuando se tiene un número grande de estratos la muestra en cada uno de ellos es pequeña y el estimador pierde precisión. Aunque no interesa efectuar inferencias sobre los estratos.

Aún cuando la precisión ganada del muestreo estratificado con respecto al MAS es importante, si la población es bastante grande y heterogénea (un país, un estado) resulta difícil contar con una lista detallada y confiable de los sujetos (aún en un estrato). El costo de la encuesta es alto por tener que visitar y aplicar el cuestionario a las personas escogidas, independientemente de qué tan lejos se encuentren una de otra. Una de las opciones que toma en cuenta costo y precisión es considerar conglomerados como unidades de muestreo, aunque se pierde precisión se espera recuperarla al estratificar. A continuación se presenta el diseño de muestreo estratificado utilizando como unidad de muestreo un conglomerado.

1.5.1 Muestreo estratificado de conglomerados con igual tamaño

1.5.1.1. Estimador de la media, varianza del estimador y estimador de su varianza

Este diseño supone una estratificación de la población y cada unidad de muestreo es un conglomerado. Para el muestreo estratificado de conglomerados de igual tamaño Sukhatme (1984, pag. 293), presenta un *estimador insesgado* de la media. En este caso $M_{hi} = M_{hj} = M_h$ para todo $i \neq j$ en el estrato h, el estimador es

$$\hat{\bar{Y}}_{st} = \frac{1}{N\bar{M}} \sum_{h=1}^{L} N_h M_h \frac{1}{n_h} \sum_{i=1}^{n_h} \bar{y}_{hi\cdot} \qquad (1.5.1.1.a)$$

en el que N conglomerados de la población han sido agrupados en L estratos, N_h es el número de conglomerados en el estrato h , n_h es el tamaño de la muestra en cada estrato con $n = \sum_{h=1}^{L} n_h$ y cada conglomerado en cada estrato, contiene M_h elementos.

Se requiere conocer $\bar{M} = \frac{\sum_{i=1}^{N} M_i}{N}$, aunque es posible utilizar en su lugar el

estimador $\bar{M}_n = \frac{\sum_{i=1}^{n} M_i}{n}$ de la muestra, pero entonces el estimador ya no es insesgado.

$\bar{Y}_{hi\cdot}$ denota a la media por elemento del *i-ésimo* conglomerado en estrato h . Con $i=1,2,\ldots, N_h$; $h=1,2,\ldots,L$.

$$\bar{Y}_{hi\cdot} = \frac{\sum_{j=1}^{M_h} Y_{hij}}{M_h} = \frac{Y_{hi}}{M_h} \qquad (1.5.1.1.b)$$

De cada estrato se seleccionan n_h conglomerados sin reemplazo y se enumeran, la varianza del estimador cuando la selección es independiente en cada estrato está dada por:

$$V(\hat{\bar{Y}}_{st}) = \frac{1}{N^2 \bar{M}^2} \sum_{h=1}^{L} N_h^2 M_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_{bh}^2 \qquad (1.5.1.1.c)$$

$$\text{con } S_{bh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{Nh}\left(\overline{Y}_{hi\cdot} - \overline{Y}_{h\cdot\cdot}\right)^2 \quad (1.5.1.1.d)$$

bajo asignación proporcional $n_h = n\frac{N_h}{N}$ la expresión para la varianza del estimador se simplifica a :

$$V(\hat{\overline{Y}}_{st})_p = \frac{N-n}{Nn}\sum_{h=1}^{L}\frac{N_h M_h^2}{N\overline{M}^2}S_{bh}^2 \quad (1.5.1.1.e)$$

Un estimador insesgado de la varianza del estimador es :

$$\hat{V}(\hat{\overline{Y}}_{st}) = \frac{1}{N^2\overline{M}^2}\sum_{h=1}^{L}N_h^2 M_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)s_{bh}^2 \quad (1.5.1.1.f)$$

$$\text{donde } s_{bh}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{nh}\left(\overline{y}_{hi\cdot} - \overline{y}_{h\cdot\cdot}\right)^2 \quad (1.5.1.1.g)$$

1.5.2 Muestreo estratificado de conglomerados con diferente tamaño

1.5.2.1 Cuatro estimadores de la media, varianza de sus estimadores y estimadores de sus varianzas

Para el caso en que los conglomerados sean de diferente tamaño, se tiene el siguiente estimador

$$\hat{\overline{Y}}_{st} = \frac{1}{N\overline{M}}\sum_{h=1}^{L}N_h\frac{\sum_{i=1}^{n_h}M_{hi}\overline{y}_{hi\cdot}}{n_h} \quad \text{estimador insesgado} \quad (1.5.2.1.a)$$

con

$$\overline{Y}_{hi\cdot} = \frac{\sum_{j=1}^{M_{hi}}Y_{hij}}{M_{hi}} = \frac{Y_{hi}}{M_{hi}} \quad (1.5.2.1.b)$$

cuando la selección es independiente en cada estrato su varianza está dada por:

$$V(\hat{\overline{Y}}_{st}) = \frac{1}{N^2\overline{M}^2}\sum_{h=1}^{L}N_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_{bh}^2 \quad (1.5.2.1.c)$$

$$\text{con } S_{bh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{Nh}(M_{hi})^2\left(\bar{Y}_{hi}.-\bar{\bar{Y}}_{h}..\right)^2 \qquad (1.5.2.1.d)$$

bajo asignación proporcional $n_h = n\dfrac{N_h}{N}$, la expresión para la varianza del estimador se simplifica a :

$$V(\hat{\bar{\bar{Y}}}_{st})_p = \frac{N-n}{Nn}\sum_{h=1}^{L}\frac{N_h}{N\bar{M}^2}S_{bh}^2 \qquad (1.5.2.1.e)$$

Un estimador insesgado de la varianza del estimador es :

$$\hat{V}(\hat{\bar{\bar{Y}}}_{st}) = \frac{1}{N^2\bar{M}^2}\sum_{h=1}^{L}N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)s_{bh}^2 \qquad (1.5.2.1.f)$$

$$\text{donde } s_{bh}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{nh}(M_{hi})^2\left(\bar{y}_{hi}.-\bar{\bar{y}}_{h}..\right)^2 \qquad (1.5.2.1.g)$$

Tres estimadores de la media pueden obtenerse si cada estimador "hereda" las características de los considerados en un MAS de conglomerados. Se ponderan en cada estrato, y son :

a) $\hat{\bar{\bar{Y}}}_{mst} = \displaystyle\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}$ *estimador media de medias por estrato* sesgado global (1.5.2.1.h)

b) $\hat{\bar{\bar{Y}}}_{ust} = \displaystyle\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{n_h\bar{M}_h}$ *estimador insesgado en cada estrato* sesgado global (1.5.2.1.i)

$$\text{donde } \bar{M}_h = \frac{\sum_{i=1}^{N_h}M_{hi}}{N_h} \qquad (1.5.2.1.j)$$

c) $\hat{\bar{\bar{Y}}}_{rst} = \displaystyle\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h}M_{hi}}$ *estimador de razón en cada estrato* sesgado global (1.5.2.1.k)

y al ser sesgados, sus respectivos errores cuadráticos medios son los siguientes

El error cuadrático medio para el estimador media de medias por estrato es

$$ECM(\hat{\bar{Y}}_{mst}) = \sum_{h=1}^{L} \frac{N_h^2}{N^2}\left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{b\,h}^2 \qquad (1.5.2.1.l)$$

$$\text{donde } S_{b\,h}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (\bar{Y}_{hi\cdot} - \bar{\bar{Y}}_h)^2 \qquad (1.5.2.1.m)$$

El estimador del error cuadrático medio es

$$\widehat{ECM}(\hat{\bar{Y}}_{mst}) = \sum_{h=1}^{L} \frac{N_h^2}{N^2}\left(\frac{1}{n_h} - \frac{1}{N_h}\right) s_{b\,h}^2 \qquad (1.5.2.1.n)$$

$$\text{donde } s_{b\,h}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (\bar{y}_{hi\cdot} - \bar{\bar{y}}_h)^2 \qquad (1.5.2.1.ñ)$$

El error cuadrático medio del estimador insesgado en cada estrato que es sesgado globalmente está dado por

$$ECM(\hat{\bar{Y}}_{ust}) = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{u\,h}^2 \qquad (1.5.2.1.o)$$

$$\text{donde } S_{u\,h}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} \left(\frac{M_{hi}\bar{Y}_{hi.}}{\bar{M}_h} - \bar{\bar{Y}}_h\right)^2 \qquad (1.5.2.1.p)$$

el estimador de su error cuadrático medio es

$$\widehat{ECM}(\hat{\bar{Y}}_{ust}) = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) s_{u\,h}^2 \qquad (1.5.2.1.q)$$

$$\text{donde } s_{u\,h}^2 = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} \left(\frac{M_{hi}\bar{y}_{hi.}}{\bar{M}_h} - \bar{\bar{y}}_h\right)^2 \qquad (1.5.2.1.r)$$

El error cuadrático medio para el estimador de razón por estrato es aproximadamente

$$ECM(\hat{\bar{Y}}_{rst}) \doteq \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{u.v\,h}^2 \qquad (1.5.2.1.s)$$

donde $S^2_{u,v_{st}} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(\bar{U}_{hi\cdot} - \bar{Y}_h V_{hi})^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\frac{M_{hi}^2}{\bar{M}_h^2}(\bar{Y}_{hi\cdot} - \bar{Y}_h)^2$ (1.5.2.1.t)

el estimador de su error cuadrático medio es

$$\widehat{ECM}(\hat{\bar{Y}}_{r_{st}}) \doteq \left(\frac{1}{n_h} - \frac{1}{N_h}\right)s^2_{u,v_h} \quad (1.5.2.1.u)$$

donde $s^2_{u,v_{st}} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(\bar{u}_{hi\cdot} - \bar{y}v_{hi})^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\frac{M_{hi}^2}{\bar{M}_h^2}(\bar{y}_{hi\cdot} - \bar{y}_h)^2$ (1.5.2.1.v)

1.5.3 Muestreo estratificado y sistemático

El muestreo estratificado y sistemático es la aplicación de dos diseños de muestreo: el muestreo estratificado, para el que se divide a la población en estratos homogéneos internamente y heterogéneos externamente y se escoge una muestra en cada estrato mediante algún diseño de muestreo con o sin asignación proporcional; y el sistemático, que considera dos casos: se genera un sólo arranque aleatorio para escoger la muestra, o se generan arranques aleatorios independientes en cada estrato (Madow y Madow, 1944).

La población $\mathcal{P}$ consiste de L estratos: $\mathcal{P}_1$, $\mathcal{P}_2$, .., $\mathcal{P}_L$ y cada estrato contiene N_h características de los elementos : $y_{h1}, y_{h2}, \ldots, y_{hN_L}$. Se quiere estimar la media de los valores $\mathcal{P}_h$ denotada por $\bar{y}$.

Se consideran dos métodos: el primero es muy utilizado por su facilidad administrativa al darle instrucciones idénticas a la gente para que seleccione muestras en lugares diferentes; el segundo implica generar, para cada estrato, un arranque aleatorio. Los procedimientos se describen a continuación.

1.5.3.1 Muestreo estratificado y sistemático con arranque aleatorio único

Si $N_h=k_h n_h$ y $k_1=k_2=...=k$, se elige al azar uno de los enteros $1,2,..,k$ cada uno con probabilidad $1/k$ de ser seleccionado. Si el entero seleccionado es i la muestra que se escoge de $\mathcal{P}_h$ consiste de los elementos y_{hi} , $y_{hi+k},...,$ $y_{hi+(n-1)k}$. Un arranque aleatorio determina el periodo para todos los estratos.

1.5.3.2. Muestreo estratificado y sistemático con arranque aleatorio independiente en cada estrato

Consiste en seleccionar al azar un número i_h de $1,2,...,$ k_h para cada h , tiene una probabilidad de $1/k_h$ de ser seleccionado. Existen exactamente k_1 , k_2 , ... ,k_L maneras de componer la muestra al realizar el procedimiento aleatorio, ya que en cada estrato la selección de la muestra es independiente. A partir del arranque aleatorio se selecciona la muestra en cada estrato, para después obtener el estimador de la media como un muestreo estatificado.

CAPÍTULO 2. MÉTODOS DE ESTIMACIÓN DE VARIANZA

2.1 Introducción

Las encuestas por muestreo se utilizan para analizar fenómenos sociales de interés, en agricultura, demografía y economía por citar algunos campos. Los beneficios que se obtienen de las encuestas por muestreo se han incrementado, también han resultado importantes los métodos para analizar e interpretar sus resultados.

Las dificultades del diseño e instrumentación de una encuesta hacen valorar lo importante de utilizar la información lo mejor posible, empleando métodos que involucren la muestra obtenida.

Para conocer la precisión de cada estimador debe estimarse su varianza, que se encuentra en función de la forma del estimador y el diseño muestral.

Para algunos diseños muestrales existe una expresión analítica para la varianza y para su estimador. Para algunos otros diseños, tan simples como el muestreo sistemático, no existe expresión analítica explícita para el cálculo de la varianza o de su estimador. En estos casos la estimación de la varianza se calcula haciendo uso de algunos métodos desarrollados. Entre los métodos de cálculo de varianza existen los de linearización de la varianza (cuando la expresión existe pero es complicada en su cálculo), métodos de remuestreo en distintas modalidades: grupos aleatorios, bootstrap, jacknife, half samples.

En este capítulo se describirá un método de remuestreo llamado grupos aleatorios, que se empleará para comparar con las fórmulas analíticas y por la facilidad de su aplicación.

2.2 El método de grupos aleatorios

Al obtener los resultados de una encuesta y el valor de sus estimadores es interesante saber si es aún posible generar alguna estimación adicional.

Uno de los resultados de la teoría del remuestreo llamada grupos aleatorios se describe a continuación. El método presenta dos maneras de aplicarse: grupos aleatorios independientes y grupos aleatorios dependientes.

Los métodos de remuestreo utilizan la muestra para realizar con ella estimaciones de los valores poblacionales.

El objetivo del método de grupos aleatorios es estimar la varianza de los estimadores, se escogen muestras de la muestra generada por un diseño de muestreo.

2.2.1 Grupos aleatorios independientes

Si se cuenta con los valores de una muestra de tamaño n, consideremos el conjunto

$$\mathcal{M} = \{ y_1, y_2, \dots, y_n \}$$

de valores de la muestra provenientes de una subpoblación

$$\mathcal{P}_{\mathcal{M}} = \{ s_1, s_2, \dots, s_n \}$$

Sea $\theta = \theta(Y_1, Y_2, \dots, Y_N)$ el parámetro de interés, sin que existan restricciones respecto a su linealidad, puede ser la razón poblacional Y sobre X, el de regresión, o el coeficiente de correlación.

Establecemos a $\mathcal{P}_{\mathcal{M}}$ como una población llamada *población fuente* y a $\mathcal{M}$ como la *muestra fuente* .

Escogemos muestras aleatorias (de acuerdo con el tipo de muestreo utilizado para obtener $\mathcal{M}$) de tamaño m $(m \leq n)$ k veces ($k \geq 2$) para formar estimaciones de θ. El procedimiento se explica a continuación.

Se toma una muestra de $\mathcal{P}_{\mathcal{M}}$ con valores $G_1 = \{ y_1, y_2, \ldots, y_m \}$ y se calcula a $\hat{\theta}_1 = \hat{\theta}_1(y_1, y_2, \ldots, y_m)$ como primer valor del estimador, las unidades de la muestra G_1 se reemplazan en $\mathcal{P}_{\mathcal{M}}$ y nuevamente se escoge otra muestra $G_2 = \{ y_1, y_2, \ldots, y_m \}$, obtenemos a $\hat{\theta}_2 = \hat{\theta}_2(y_1, y_2, \ldots, y_m)$ como segundo valor del estimador. Las unidades de G_2 se reemplazan en la población fuente y el proceso se repite hasta obtener k muestras y k valores del estimador .

A las k muestras obtenidas se les denominará *grupos aleatorios independientes*, ya que la selección de cada una de ellas se hace con reemplazo y por lo tanto son independientes.

Al generar k muestras o grupos aleatorios es posible obtener otro estimador como combinación de los k valores, además se puede investigar qué propiedades tiene este nuevo estimador y calcular su varianza.

Mediante el siguiente teorema se establecen las características de este estimador.

Teorema 2.1 (Wolter, 1985, pag. 21)

Sean $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ variables aleatorias no correlacionadas con valor esperado $E(\hat{\theta}_\alpha) = \theta$ $\alpha = 1, 2, \ldots, k$. Si $\bar{\hat{\theta}}$ se define como:

$$\bar{\hat{\theta}} = \frac{\sum_{\alpha=1}^{k} \hat{\theta}_\alpha}{k} \qquad (2.2.1.a)$$

entonces

1) $E(\bar{\hat{\theta}}) = \theta$, $\bar{\hat{\theta}}$ es un estimador insesgado de θ (2.2.1.b)

$$2)\ v(\bar{\hat{\theta}}) = \frac{\sum_{\alpha=1}^{k}(\hat{\theta}_\alpha - \bar{\hat{\theta}})^2}{k(k-1)} \text{ es un estimador insesgado de Var } \{\bar{\hat{\theta}}\} \qquad (2.2.1.c)$$

El método de grupos aleatorios fue uno de los primeros procedimientos desarrollados para simplificar la estimación de varianza en encuestas por muestreo complejas.

La aplicación del método la realizó Mahalanobis en encuestas de producción agrícola en Bengal, las llamó *muestras interpenetrantes.*

Los resultados de aplicar el método en muestreo estadístico fueron presentados por Mahalanobis al rendir un informe referente al periodo 1937-1945 en el Instituto de Estadística de la India (Mahalanobis, 1946).

Mahalanobis utiliza submuestras en su análisis estadístico por considerarlo conveniente y práctico, asigna a las unidades de muestreo un número consecutivo en el orden en que fueron escogidas para estar en la muestra . Divide los datos de la muestra en un número de conjuntos aleatorios independientes, escoge unidades con número par o impar o con números terminados en algunos dígitos en particular.

Los trabajos de cálculo y análisis se hacen separadamente para cada subconjunto, esta manera de proceder tiene tres ventajas: primero, los resultados se obtienen muy rápido; segundo, al comparar los resultados de los diferentes subconjuntos se pueden detectar errores graves en el procedimiento o en los cálculos y finalmente, los resultados suministran una buena idea del margen de error efectivo o general.

Además de Mahalanobis otros autores se han referido al teorema 2.1, aunque mencionan condiciones más fuertes para las variables al pedir que sean independientes e idénticamente distribuidas (Des Raj, 1980, pag. 52), o sólo independientes (Sukhatme, 1984, pag. 18), bajo las que se aplica el resultado.

El método de grupos aleatorios puede usarse como alternativa a la fórmula de la estimación de varianza empleada habitualmente, porque no genera dificultad adicional alguna o costos significantes. Incluso, es posible obtener reducciones de trabajo y costo, se puede estimar la varianza de no sólo uno, sino de un número considerable de estadísticos diferentes. La reducción puede ser muy sustancial (Hansen et al., 1953, pag. 440).

Se empleará el término grupo aleatorio independiente cuando se haga referencia a este método de estimación de varianza.

El estadístico $\bar{\hat{\theta}}$ puede usarse como un estimador de θ, mientras que $v(\bar{\hat{\theta}})$ es el estimador de grupo aleatorio para su varianza, Var $\{\bar{\hat{\theta}}\}$.

Aunque el teorema requiere que las variables aleatorias $\hat{\theta}_\alpha$ sean no-correlacionadas, el procedimiento de reemplazar la muestra $G_{\alpha-1}$ antes de seleccionar la muestra G_α tiende a inducir independencia entre las $\hat{\theta}_\alpha$. El método de muestreo de grupos aleatorios independientes satisface los requerimientos del Teorema.

Las inferencias acerca del parámetro θ se basan, generalmente, en la teoría de la Normal o la teoría de t de Student, el resultado se establece en el siguiente teorema.

Teorema 2.2 (Wolter, 1985, pag. 22).

Sean $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ variables aleatorias, independientes e idénticamente distribuidas como una Normal (θ, σ^2). Entonces

1) El estadístico $z = \dfrac{(\bar{\hat{\theta}} - \theta)}{\sqrt{\dfrac{\sigma^2}{k}}}$ se distribuye como una Normal $(0,1)$ (2.2.1.d)

2) el estadístico $t = \dfrac{(\bar{\hat{\theta}} - \theta)}{\sqrt{v(\bar{\hat{\theta}})}}$ se distribuye como una t de Sudent,

con k-1 grados de libertad. (2.2.1.e)

3) Si la varianza de $\bar{\hat{\theta}}$ es conocida o si k es muy grande ($k \geq 25$), entonces un intervalo de confianza al $(1$-$\alpha)$ % para θ es :

$$(\bar{\hat{\theta}} - z_{\alpha/2}\sqrt{v(\bar{\hat{\theta}})}\ ,\ \bar{\hat{\theta}} + z_{\alpha/2}\sqrt{v(\bar{\hat{\theta}})}) \qquad (2.2.1.f)$$

donde $z_{\alpha/2}$ es el percentil de una distribución normal

4) Si la varianza de $\bar{\hat{\theta}}$ se desconoce y k no es grande, el intervalo de confianza para $\bar{\hat{\theta}}$ es

$$(\bar{\hat{\theta}} - t_{k-1,\alpha/2}\sqrt{v(\bar{\hat{\theta}})}\ ,\ \bar{\hat{\theta}} + t_{k-1,\alpha/2}\sqrt{v(\bar{\hat{\theta}})}) \qquad (2.2.1.g)$$

donde $t_{k-1,\alpha/2}$ es el valor del percentil de una distribución t de Student con k-1 g.l.

De manera semejante, las pruebas de hipótesis acerca de θ pueden basarse en el teorema.

Es importante mencionar que el estimador de varianza no sólo estima la varianza del promedio del estimador, sino también a la varianza del estimador mismo es decir, $v(\bar{\hat{\theta}}) = v(\hat{\theta})$ (Wolter, 1985, pag. 24).

2.2.2 Grupos aleatorios dependientes

En algunos casos, las muestras se remuestrean utilizando MASSR en lugar de seleccionarlas como grupos aleatorios independientes. Los grupos aleatorios se forman con una selección aleatoria al dividir ó particionar la muestra fuente en k grupos como se explica más adelante.

Los estimadores se calculan para cada grupo aleatorio y la varianza se estima al usar una expresión de la forma presentada en el Teorema 2.1

Los estimadores $\hat{\theta}_\alpha$ del grupo aleatorio están correlacionados porque el muestreo se realiza sin reemplazo y el resultado del Teorema. 2.1.1 no es necesariamente cierto, el estimador de grupo aleatorio tiene sesgo (Wolter, 1985, pag. 30).

El método de *grupos aleatorios dependientes* consiste en escoger k muestras sin reemplazo de la población fuente $\mathcal{P}_\mathcal{M}$. El procedimiento se explica a continuación.

Si $\mathcal{M} = \{y_1, y_2, \dots, y_n\}$ es el conjunto de valores de la muestra fuente, entonces $G_\alpha = \{y_1, y_2, \dots, y_m\}$ con $\alpha = 1, 2, .. k$ es el *k-ésimo* grupo aleatorio dependiente de tamaño m, pero con muestras escogidas sin reemplazo.

La formación de los grupos aleatorios depende del tipo de muestreo utilizado en la encuesta. Si el tipo de muestreo es *muestreo aleatorio simple*, los grupos aleatorios podrán formarse al dividir la población fuente aleatoriamente. El primer grupo se obtiene al seleccionar una m.a.s.s.r. de tamaño $m=n/k$ de la muestra fuente,

$$G_1 = \{y_1, y_2, \dots, y_{n/k}\}$$

Después de haberlas escogido, las m unidades seleccionadas se separan de la población fuente, ya que se utiliza un muestreo sin reemplazo.

El segundo grupo aleatorio, es de nueva cuenta una muestra de tamaño m de las $n - m$ unidades restantes, es decir, de las $n - m = n - n/k = n(1-1/k)$ unidades que quedan después de separar las unidades seleccionadas.

Si n/k no es un entero es decir, $n=km+q$ con $0<q<k$ y $n/k = m + q/k = m + r$ (es un entero más un residuo), las q unidades que sobran pueden dejarse fuera del *k-ésimo* grupo o

agregarlas a cada uno de los primeros q grupos, el número de grupos aleatorios está condicionado a que existan unidades de donde sea posible tomar una muestra.

Si la muestra se elige de un *muestreo sistemático* con un arranque aleatorio, los grupos aleatorios podrán formarse al dividir la muestra fuente sistemáticamente. Esto es posible hacerlo al generar aleatoriamente un entero entre 1 y k digamos α^*, la primera unidad de la muestra fuente se asigna al grupo aleatorio
α-*ésimo*, la segunda al grupo aleatorio α^*+1 y así sucesivamente .

En un *muestreo multietápico* los grupos aleatorios pueden formarse al dividir en k grupos los agrupamientos finales, es decir, la unión de todas las unidades elementales seleccionadas de la misma unidad primaria de muestreo. Todas las unidades de segunda, tercera y sucesivas etapas provenientes de las seleccionadas de unidades primarias de muestreo deben tratarse como una sola unidad simple cuando se forman los grupos aleatorios.

Por ejemplo, si se toma al estado de Guanajuato y se supone un muestreo bietápico, la unidad de primera etapa es el municipio y la de segunda etapa es la sección.

Se tienen 46 municipios y 3005 secciones, se eligen al azar n municipios, y de cada uno de ellos se toma una muestra de tamaño m_i con i=1,2,...,n para obtener la muestra

$$\mathcal{M}=\left\{\left\{y_{11},y_{12}\cdots,y_{1m_1}\right\},\left\{y_{21},y_{22}\cdots,y_{2m_2}\right\},\cdots,\left\{y_{n1},y_{n2}\cdots,y_{nm_n}\right\}\right\}$$

donde y_{ij} es la votación para un partido en la sección j del municipio i de la muestra i=1,2,...n ; j=1,2,...,m_n , se tienen n agrupamientos.

Para formar los grupos aleatorios se toma a manera de ejemplo los grupos independientes y k=4.

El primer grupo se forma de la siguiente manera:

1) Se elige al azar *n'* de los *n* agrupamientos de la muestra, de los elegidos se toma una muestra sin reemplazo (por ser grupos aleatorios dependientes) de tamaño $m_i' = \frac{m_i}{4}$ para tener como primer grupo aleatorio a

$$G_1 = \left\{\{y_{11}, y_{12} \ldots, y_{1m_1'}\}, \{y_{21}, y_{22} \ldots, y_{2m_2'}\}, \ldots, \{y_{n'1}, y_{n'2} \ldots, y_{n'm_{n'}'}\}\right\}$$

los demás grupos se forman de manera análoga al hacer la selección de los elementos no seleccionados en el grupo anterior , se obtienen entonces 4 valores del estimador

$$\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \quad \bar{\hat{\theta}} = \frac{\sum_{\alpha=1}^{k} \hat{\theta}_\alpha}{k} \quad \text{y} \quad v(\bar{\hat{\theta}}) = \frac{\sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \bar{\hat{\theta}})^2}{k(k-1)}$$

Las selección en forma aleatoria simple o sistemática de las unidades elementales depende del tipo de muestreo realizado en la primera etapa.

En el *muestreo estratificado* pueden escogerse dos opciones: uno, se puede estimar la varianza dentro de un estrato al utilizar las reglas ya enunciadas y según el tipo de muestreo utilizado; dos, se puede estimar la varianza total.

Por ejemplo, si Guanajuato se estratifica en 18 distritos electorales se puede estimar la varianza del estimador con grupos aleatorios respetando el diseño de muestreo en cada estrato, es decir, si el muestreo en el estrato es un MAS o un muestreo sistemático se utilizan las reglas enunciadas. Cada estrato tendrá su estimador de varianza.

Para estimar la varianza total entre todos los estratos, cada grupo aleatorio debe ser una muestra estratificada e incluir unidades de cada estrato. El primer grupo aleatorio se obtiene al seleccionar una m.a.s.s.r. de tamaño $m_h = n_h / 4$ de la población fuente $\mathcal{P}_h$ por cada estrato h con $h=1,2,\ldots,18$. El segundo se forma de manera análoga para cada estrato al seleccionar una muestra de las $n_h - m_h$ unidades restantes, se repite el proceso hasta que sea posible y se obtienen 4 grupos aleatorios, que son una muestra estratificada con la que se evalúa el estimador. Después de obtener 4 valores del estimador, se promedian para calcular el estimador de varianza.

Un *procedimiento equivalente* se utilizará para obtener los grupos aleatorios dependientes, consiste en hacer una partición de los valores de la muestra en *k* grupos como se explica a continuación.

Si $k=2$ los elementos de la muestra se separan en pares e impares; cuando $k=3$ el primer elemento se envia al grupo 1, el segundo al grupo 2 y el tercero al tercer grupo, el cuarto va al grupo 1, el quinto al 2 y así sucesivamente hasta conformar la muestra.

Es posible utilizar el método de grupos aleatorios en forma repetida para diseños estratificados con igual proporción de K unidades por estrato (Kovar et al, 1988). A continuación se explica la manera de aplicarlo.

La muestra es aleatoriamente dividida en K submuestras aleatorias sin reemplazo en cada estrato ($m_i = n_i/K$), se calculan sus medias para ponderarlas y obtener $\hat{\theta}_\alpha$ estimaciones de θ ($\alpha = 1,2,\ldots,K$.)

El procedimiento se repite R veces para obtener dos estimadores de la varianza, que son

$$v_1 = \frac{\sum_{r=1}^{R}\{\sum_{\alpha=1}^{K}(\hat{\theta}_{r\alpha} - \hat{\theta})^2 / K(K-1)\}}{R}$$ donde $\hat{\theta}$ es el estimador del parámetro según el diseño de muestreo con toda la muestra original (2.2.2.a)

y

$$v_2 = \frac{\sum_{r=1}^{R}\{\sum_{\alpha=1}^{K}(\hat{\theta}_{r\alpha} - \overline{\theta}_r)^2 / K(K-1)\}}{R} \quad \text{donde} \quad \overline{\theta}_r = \sum_{\alpha=1}^{K}\hat{\theta}_{r\alpha} / K. \qquad (2.2.2.b)$$

Una observación que se desprende, es que los dos estimadores pueden obtenerse para cualquier diseño de muestreo y con sólo una muestra.

El método de grupos aleatorios independientes utiliza muestras de tamaño m; sin embargo, no menciona cuál debe ser su tamaño. Para estimar la varianza en este trabajo se utilizó m proporcional al número de grupos aleatorios, es decir, $m = n / k$.

CAPÍTULO 3. APLICACIONES A CONTEOS RÁPIDOS

3.1 Introducción

En este capítulo se describe primeramente el concepto de conteo rápido, se presenta la manera de realizar simulaciones de conteos rápidos con resultados electorales de tres estados de la República Mexicana.

Los diseños muestrales que se utilizan son: MAS, muestreo sistemático, y muestreo estratificado, se calculan los estimadores de la proporción de votos y se estima su varianza mediante grupos aleatorios independientes y dependientes.

3.2 Conteo rápido

Un conteo rápido es el procedimiento que el día de la elección se aplica a secciones electorales elegidas al azar bajo un diseño de muestreo, y que consiste en recabar la información de los votos emitidos al cerrarse la sección oficialmente.

El interés principal en un conteo rápido es estimar el porcentaje de votos válidos a favor de los candidatos de los partidos mayoritarios contendientes en una elección.

La información del conteo rápido debe comunicarse con la mayor rapidez posible, para procesarla, y obtener una estimación confiable y oportuna.

Mediante simulaciones de conteos rápidos se investigan las propiedades de algunos diseños muestrales. Se utilizan resultados oficiales de las elecciones para gobernador efectuadas en los estados de Guanajuato, Yucatán y Michoacán en el año de 1995.

Con la información de los resultados oficiales por casilla de cada estado se conocen los valores poblacionales de la proporción de votos para cada partido y su varianza. También se calculan los estimadores de varianza de proporciones estimadas bajo algunos diseños de interés.

Se obtuvieron resultados de simulación al considerar los siguientes diseños de muestreo: MAS, muestreo sistemático (después de obtener la muestra en forma sistemática se le considera como una muestra aleatoria simple para la evaluación de los estimadores y de su varianza) y el muestreo estratificado con algunas variantes.

Las unidades de muestreo utilizadas fueron conglomerados de electores llamados secciones excepto en Guanajuato, que para su estratificación rural - urbana se utilizaron a las casillas como unidad de muestreo.

A continuación se presenta la estructura de las bases de datos y se describe la notación utilizada para cada diseño muestral.

La característica de interés (voto en favor de un partido) es una variable que particularmente sólo toma los valores uno o cero para reducirse a una proporción.

3.3 Descripción de las bases de datos, de los estimadores considerados y del cálculo de varianza por grupos aleatorios

Las bases de datos tienen en general los siguientes campos :

Campo	Nombre_campo	tipo	rango
1	DLOCAL	Numérico	6
2	CASILLA	Numérico	10
3	TIPO	Alfanumérico	2

4	SECCION[1]	Numérico	10
5	PAN	Numérico	4
6	PRI	Numérico	4
7	PRD	Numérico	8
8	OTROS	Numérico	10
9	VALIDOS	Numérico	8
10	UBICACION	Alfanumérico	10
11	LIST_NOM	Numérico	10
12	NULOS	Numérico	10
13	NO_REG	Numérico	10

[1] Los campos se escribirán sin acento

DLOCAL contiene el número de distrito electoral.

CASILLA es el número de casilla dado por el IFE local.

TIPO es el tipo de casilla: básica, contigua, especial, o extraordinaria.

SECCION es el número de la sección electoral dado por el IFE local.

PAN es el total de votos válidos para el Partido de Acción Nacional por casilla o sección.

PRI es el total de votos válidos para el Partido Revolucionario Institucional por casilla o sección.

PRD es el total de votos válidos para el Partido de la Revolución Democrática por casilla o sección.

OTROS es el total de votos válidos para los demás partidos contendientes como el PFCRN, PT, PVEM y en algunos casos como los estados de Guanajuato y Yucatán incluye al PRD, ya que no alcanza un porcentaje de votación superior a los partidos mencionados.

VALIDOS es el total de votos válidos para la casilla o sección.

UBICACION es la característica urbana o rural de la casilla.

LISTA_NOM es el número de electores registrados en la lista nominal el día de la elección en la casilla o sección.

NULOS votos anulados en la casilla o sección.

NO_REG votos a candidatos no registrados en la casilla o sección.

Cada registro corresponde a una casilla o sección, incluye la votación para los diferentes partidos. Un cuadro descriptivo se presenta con la información de las bases de datos: número de secciones, lista nominal, porcentaje de participación, votos válidos, promedio de votos válidos, máximo y mínimo de votos válidos, máximo y mínimo de lista nominal y promedio de lista nominal.

DESCRIPCIÓN	GUANAJUATO	YUCATÁN	MICHOACÁN *
NÚMERO DE SECCIONES	3,005	1,054	2,081
LISTA NOMINAL	2,163,111	748,335	1,474,589
PORCENTAJE DE PARTICIPACIÓN	57.55	67.23	60.14
VOTOS VÁLIDOS	1,244,955	503,137	886,794
PROMEDIO VOTOS VÁLIDOS	414	477	426
MÁXIMO DE VOTOS VÁLIDOS	1,190	1,435	1,670
MÁXIMO DE LISTA NOMINAL	1,874	1,965	1,815
MÍNIMO DE VOTOS VÁLIDOS	15	32	23
MÍNIMO DE LISTA NOMINAL	35	51	45

TABLA 1 Resumen de la información contenida en las tres archivos fuente (por secciones)

* <u>Nota aclaratoria</u> : La base de datos disponible para el estado de Michoacán incluye únicamente 13 distritos electorales. Las simulaciones se realizaron con esta información, aunque de acuerdo al Instituto Federal Electoral local los distritos electorales son 18.

En la siguiente parte se expone el modelo teórico en primer término, para después relacionarlo con la información de las bases de datos disponibles por cada uno de los estados y según el diseño de muestreo.

3.3.1 Muestreo aleatorio simple de conglomerados

3.3.1.1 Estimación de la media, tres estimadores

Para este diseño la unidad de muestreo es la sección electoral, la población está compuesta por N conglomerados de diferente tamaño

$$\wp = \{s_1, s_2, \ldots, s_N\}$$

donde s_i es la *i-ésima* sección electoral con voto válido, $i=1,2,\ldots,N$; N es el número de secciones, y cada sección se puede definir como:

$$s_i = \{e_{i1}, e_{i2}, \ldots, e_{ij}, \ldots, e_{iM_i}\}$$

e_{ij} es el *j-ésimo* elector de la sección *i-ésima* con $j=1,2,\ldots, M_i$ y M_i es el número de electores en la sección i.

Al evaluar la variable de interés Y que mide la preferencia electoral por un determinado partido se obtiene que

$$Y(s_i) = \{Y_{i1}, Y_{i2}, \ldots, Y_{ij}, \ldots, Y_{iM_i}\}$$

donde Y_{ij} es uno si el *j-ésimo* elector de la sección *i-ésima* vota por un determinado partido, y cero si no es así.

Es posible definir una variable en términos de estos valores, que permita simplificar los cálculos de la siguiente manera.

$Y_i = \sum_{j=1}^{M_i} Y_{ij}$ representa el número de votos válidos para un determinado partido en la sección *i-ésima*, podemos definir a los valores de la población agrupados en conglomerados como sigue.

$$\wp' = \{Y_1, Y_2, \ldots, Y_N\}.$$

En la base de datos Y_i representa al registro *i-ésimo* del campo correspondiente al partido en cuestión; se elige una m.a.s.s.r. de tamaño n y se evalúan los tres estimadores siguientes, *i*) $\hat{\bar{Y}} = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$ estimador media de medias,

ii) $\hat{\bar{Y}}_u = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}}$ estimador insesgado y

iii) $\hat{\bar{Y}}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$ estimador de razón con $\bar{y}_{i\cdot} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i} = \frac{y_i}{M_i}$ $i=1,..,n$

3.3.1.2 Estimación de la varianza, dos estimadores

Se estimó la varianza de los estimadores media de medias y de razón que sólo utilizan información de la muestra.

Es deseable que los grupos aleatorios sean lo suficientemente grandes y que se acerquen al tamaño de la muestra total. Por esta razón se eligió k=2 y 4 .

Las muestras generadas para calcular los estimadores en un MAS se utilizan para formar grupos aleatorios independientes y dependientes con k=2 y 4 de la siguiente manera.

Se denota $\mathcal{M}$ como la muestra fuente de tamaño n

$$\mathcal{M} = \{ y_1, y_2, \ldots, y_n \}$$

provenientes de una subpoblación

$$\mathcal{P}_{\mathcal{M}} = \{ s_1, s_2, \ldots, s_n \}$$

de la población de secciones y se escogen muestras para estimar la varianza de

$$\hat{\bar{Y}} = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n} \quad \text{y} \quad \hat{\bar{Y}}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$$

Como son sesgados, se calcula su error cuadrático medio por las fórmulas (1.3.2.1.f) y (1.3.2.1.n).

3.3.1.2.1 Grupos aleatorios independientes con k=2,4

Para obtener los grupos aleatorios independientes se considera el tamaño de las muestras proporcional al de la muestra fuente, $m=n/k$ o aproximadamente con $k=2,4$ donde m es el número de elementos del grupo aleatorio.

El primer grupo aleatorio está compuesto por una m.a.s.s.r. de $\mathcal{P}_M$ de tamaño $m=n/k$, los valores que se obtienen son :

$$G_1 = \{y_1, y_2, \ldots, y_m\}$$

donde cada y_i es la suma de votos favorables a un partido en la muestra. Las unidades escogidas se reemplazan en la muestra fuente, ya que son grupos independientes y nuevamente se escoge una muestra de tamaño n/k para obtener el segundo grupo aleatorio:

$$G_2 = \{y_1, y_2, \ldots, y_m\}$$

Se continúa el proceso hasta obtener 2 o 4 grupos aleatorios, para los dos estimadores se procede así:

Para el estimador media de medias $\hat{\theta} = \hat{\bar{Y}} = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$ se evalúan los valores de los k grupos aleatorios ($k=2,4$) es decir,

$$\hat{\theta}_1 = \hat{\theta}_1(y_1, y_2, \ldots, y_m) = \frac{\sum_{i=1}^{m} \bar{y}_{i\cdot}}{m}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_1, y_2, \ldots, y_m) = \frac{\sum_{i=1}^{m} \bar{y}_{i\cdot}}{m}$$

$$\hat{\theta}_k = \hat{\theta}_k(y_1, y_2, \ldots, y_m) = \frac{\sum_{i=1}^{m} \bar{y}_{i\cdot}}{m}, \quad \text{con } k=2,\ 4$$

para el caso del estimador de razón $\hat{\theta} = \hat{\bar{Y}}_r = \dfrac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$ se procede de manera análoga a la del estimador anterior,

$$\hat{\theta}_1 = \hat{\theta}_1(y_1, y_2, \ldots, y_m) = \frac{\sum_{i=1}^{m} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{m} M_i}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_1, y_2, \ldots, y_m) = \frac{\sum_{i=1}^{m} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{m} M_i}$$

.

.

.

$$\hat{\theta}_k = \hat{\theta}_k(y_1, y_2, \ldots, y_m) = \frac{\sum_{i=1}^{m} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{m} M_i}$$

con $k=2,\ 4$

Sus valores se promedian para obtener el valor $\bar{\hat{\theta}} = \dfrac{\sum_{\alpha=1}^{k} \hat{\theta}_\alpha}{k}$ y el estimador de varianza

$v(\bar{\hat{\theta}}) = \dfrac{\sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \bar{\hat{\theta}})^2}{k(k-1)}$ para cada estimador.

Debe recordarse que el estimador de varianza no sólo estima la varianza del promedio del estimador, sino también a la varianza del estimador mismo (Wolter, 1985, pag. 24)

3.3.1.2.2 Grupos aleatorios dependientes con k=2,4

Para los grupos aleatorios dependientes se utiliza el procedimiento equivalente (sección 2.2.2.) para tomar muestras sin reemplazo. Con los valores de la muestra obtenida $\mathcal{M} = \{y_1, y_2, \ldots, y_n\}$ se hace una partición en k grupos aleatorios dependientes con $k = 2, 4$.

Si $k=2$ se forman dos grupos, una de las maneras se describe enseguida:
El primer grupo contiene a los elementos impares y el segundo a los elementos pares:

$$G_1 = \{y_1, y_3, \ldots, y_z\}$$
$$G_2 = \{y_2, y_4, \ldots, y_{z'}\}$$

si n es par, $z=n-1$ y $z'=n$; si n es impar, $z=n$ y $z'=n-1$.

Si $k=4$ es posible integrarlos del siguiente modo, se particiona a la muestra fuente en cuatro grupos,

$$G_1 = \{y_1, y_5, \ldots, y_a\}$$
$$G_2 = \{y_2, y_6, \ldots, y_b\}$$
$$G_3 = \{y_3, y_7, \ldots, y_c\}$$
$$G_4 = \{y_4, y_8, \ldots, y_d\} \text{ con } a,b,c,d \leq \frac{n}{4}+1$$

una vez formados los grupos aleatorios, se evalúan los estimadores

Para $\hat{\theta} = \hat{\bar{Y}} = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$ el estimador media de medias, se evalúan los valores de los 4 grupos aleatorios (para k=2, es análogo) es decir,

$$\hat{\theta}_1 = \hat{\theta}_1(y_1, y_2, \ldots, y_a) = \frac{\sum_{i=1}^{a} \bar{y}_{i\cdot}}{a}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_1, y_2, \ldots, y_b) = \frac{\sum_{i=1}^{b} \bar{y}_{i\cdot}}{b}$$

$$\hat{\theta}_3 = \hat{\theta}_3(y_1, y_2, \ldots, y_c) = \frac{\sum_{i=1}^{c} \bar{y}_{i\cdot}}{c}$$

$$\hat{\theta}_4 = \hat{\theta}_4(y_1, y_2, \ldots, y_d) = \frac{\sum_{i=1}^{d} \bar{y}_{i\cdot}}{d}$$

para el caso del estimador de razón $\hat{\theta} = \hat{\bar{Y}}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$ se procede de manera análoga al estimador anterior.

$$\hat{\theta}_1 = \hat{\theta}_1(y_1, y_2, \ldots, y_a) = \frac{\sum_{i=1}^{a} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{a} M_i}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_1, y_2, \ldots, y_b) = \frac{\sum_{i=1}^{b} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{b} M_i}$$

$$\hat{\theta}_3 = \hat{\theta}_3(y_1, y_2, \ldots, y_c) = \frac{\sum_{i=1}^{c} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{c} M_i}$$

$$\hat{\theta}_4 = \hat{\theta}_4(y_1, y_2, \ldots, y_d) = \frac{\sum_{i=1}^{d} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{d} M_i}$$

se promedian para obtener el valor $\bar{\hat{\theta}} = \frac{\sum_{\alpha=1}^{k} \hat{\theta}_\alpha}{k}$ y el estimador de varianza

$v(\bar{\hat{\theta}}) = \frac{\sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \bar{\hat{\theta}})^2}{k(k-1)}$ para cada uno de ellos.

3.3.2 Muestreo sistemático de conglomerados

Se calcula a $k=N/n$ y se elige al azar un número i entre 1 y k para que la muestra se forme con los elementos i, $i+k$, $i+2k$, ... , $i+(n-1)k$. El estimador de la media corresponde a la media muestral, sin embargo no existe fórmula explícita para la varianza por lo que para fines de cálculo de varianza se considera una muestra aleatoria simple para evaluar los estimadores y, para evaluar a su varianza se considera al estimador correspondiente a un MAS.

La estimación de varianza mediante grupos aleatorios independientes y dependientes se realiza de la misma manera que con una muestra aleatoria simple.

3.3.3 Muestreo estratificado de conglomerados

Para este tipo de muestreo son cuatro los estimadores de la media (se describieron en la sección 1.5.2.1), además se generan estimaciones de la varianza para dos de ellos.

3.3.3.1 Estimación de la media, cuatro estimadores

La muestra se elige en forma independiente y mediante MASSR con asignación proporcional en cada estrato, se calcula el estimador de la proporción y se pondera por el peso de cada estrato para obtener el estimador del parámetro correspondiente al diseño.

La población estratificada está compuesta por

$$\wp = \{\{s_{11}, s_{12}, \ldots, s_{1N_1}\}, \{s_{21}, s_{22}, \ldots, s_{2N_2}\}, \ldots, \{s_{L1}, s_{L2}, \ldots, s_{LN_L}\}\}$$

donde s_{hi} es una sección de electores con voto válido $h=1,2,\ldots,L$; $i=1,2,\ldots,N_h$ y N_h es el número de secciones en el estrato h.

A cada sección se le puede definir como :

$$s_{hi} = \{e_{hi1}, e_{hi2}, \ldots, e_{hij}, \ldots, e_{hiM_{hi}}\}$$

e_{hij} es el *j-ésimo* elector de la sección *i-ésima* del estrato h , $j=1,2,\ldots, M_{hi}$. M_{hi} es el número de electores de la sección i en el estrato h con $i=1,2,\ldots, N_h$.

Al medir la preferencia electoral hacia un determinado partido, se utiliza la variable Y

$$Y(s_{hi}) = \{Y_{hi1}, Y_{hi2}, \ldots, Y_{hij}, \ldots, Y_{hiM_{hi}}\}$$

donde Y_{hij} es uno si el elector *j-ésimo* de la sección *i-ésima* del *h-ésimo* estrato vota por un determinado partido, y cero si no es así.

Es posible definir una variable en términos de estos valores:

$Y_{hi} = \sum_{j=1}^{M_{hi}} Y_{hij}$ que representa el número de votos para un determinado partido en la sección *i-ésima* del estrato *h-ésimo* los valores de la población se pueden describir como :

$$\wp` = \{\{Y_{11}, Y_{12}, \ldots, Y_{1N_1}\}, \{Y_{21}, Y_{22}, \ldots, Y_{2N_2}\}, \ldots, \{Y_{L1}, Y_{L2}, \ldots, Y_{LN_L}\}\}$$

Se calculan los cuatro estimadores descritos en las sección 1.5.2.1.

i) $\hat{\bar{Y}}_{st} = \frac{1}{N\bar{M}} \sum_{h=1}^{L} N_h \frac{1}{n_h} \sum_{i=1}^{n_h} M_{hi} \bar{y}_{hi\cdot}$ estimador insesgado, *ii)* $\hat{\bar{Y}}_{mst} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} \bar{y}_{hi\cdot}}{n_h}$ estimador media de medias, *iii)* $\hat{\bar{Y}}_{ust} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} M_{hi} \bar{y}_{hi\cdot}}{n_h \bar{M}_h}$ estimador insesgado en cada estrato con

$$\overline{M}_h = \frac{\sum_{i=1}^{N_h} M_{hi}}{N_h} \text{ y } \hat{\bar{Y}}_{rst} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$$ estimador de razón en cada estrato con

$$\bar{y}_{hi\cdot} = \frac{\sum_{j=1}^{M_{hi}} y_{hij}}{M_{hi}} = \frac{y_{hi}}{M_{hi}}$$

3.3.3.2 Estimación de la varianza

La estimación de la varianza de los estimadores en un muestreo estratificado utiliza las muestras generadas por cada simulación. Se forman grupos aleatorios independientes y dependientes para $k=2,4$ de la siguiente manera .

Si consideramos a $\mathcal{M}$ como los valores de la muestra obtenida de tamaño n con asignación proporcional

$$\mathcal{M} = \{\{y_{11}, y_{12}, \ldots, y_{1n_1}\}, \{y_{21}, y_{22}, \ldots, y_{2n_2}\}, \ldots, \{y_{L1}, y_{L2}, \ldots, y_{Ln_L}\}\}$$

Como la generación de grupos aleatorios debe respetar el diseño de muestreo, cada uno debe formarse utilizando la muestra estratificada. Se calcularon sólo para los estimadores cuyos valores dependen exclusivamente de la muestra como son :

$$\hat{\bar{Y}}_{mst} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} \bar{y}_{hi\cdot}}{n_h} \text{ y } \hat{\bar{Y}}_{rst} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$$

sus errores cuadráticos medios se calculan mediante las fórmulas descritas en las secciones (1.5.2.1.*l*) y (1.5.2.1.*s*).

3.3.3.2.1 Grupos aleatorios independientes con k=2,4

Cuando se forman los grupos aleatorios independientes se toman muestras de la muestra estratificada. Se forman k grupos aleatorios, el primero consiste de muestras aleatorias sin reemplazo de cada estrato de tamaño $m_i=n_i/k$ (con tamaños proporcionales) , el primer grupo es :

$$G_1 = \{\{y_{11}, y_{12}, \ldots, y_{1m_1}\}, \{y_{21}, y_{22}, \ldots, y_{2m_2}\}, \ldots, \{y_{L1}, y_{L2}, \ldots, y_{Lm_L}\}\}$$

el segundo se genera de manera análoga

$$G_2 = \{\{y_{11}, y_{12}, \ldots, y_{1m_1}\}, \{y_{21}, y_{22}, \ldots, y_{2m_2}\}, \ldots, \{y_{L1}, y_{L2}, \ldots, y_{Lm_L}\}\}$$ se continúa hasta obtener k grupos.

El *k-ésimo* grupo es :

$$G_k = \{\{y_{11}, y_{12}, \ldots, y_{1m_1}\}, \{y_{21}, y_{22}, \ldots, y_{2m_2}\}, \ldots, \{y_{L1}, y_{L2}, \ldots, y_{Lm_L}\}\}$$

Los dos estimadores se evalúan y ponderan por el peso de cada estrato.

Para $\hat{\theta} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{m_h} \overline{y}_{hi\cdot}}{n_h}$ el estimador media de medias por estrato

$$\hat{\theta}_1 = \hat{\theta}_1(y_{11}, y_{12}, \ldots, y_{1m_1}, y_{21}, y_{22}, \ldots, y_{2m_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lm_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{m_h} \overline{y}_{hi\cdot}}{m_h}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_{11}, y_{12}, \ldots, y_{1m_1}, y_{21}, y_{22}, \ldots, y_{2m_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lm_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{m_h} \overline{y}_{hi\cdot}}{m_h}$$

.

.

.

$$\hat{\theta}_k = \hat{\theta}_k(y_{11}, y_{12}, \ldots, y_{1m_1}, y_{21}, y_{22}, \ldots, y_{2m_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lm_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{m_h} \overline{y}_{hi\cdot}}{m_h}$$

se promedian para obtener el valor $\overline{\hat{\theta}} = \frac{\sum_{\alpha=1}^{k} \hat{\theta}_\alpha}{k}$ y el estimador de varianza

$$v(\bar{\hat{\theta}}) = \frac{\sum_{\alpha=1}^{k}(\hat{\theta}_\alpha - \bar{\hat{\theta}})^2}{k(k-1)} \quad \text{para cada estimador.}$$

Para el estimador $\hat{\bar{Y}}_{rm} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$ se procede de manera similar .

3.3.3.2.2 Grupos aleatorios dependientes con k=2,4

Los valores de la muestra obtenida se separan de acuerdo al caso $k=2,4$. Si $k=2$, el primer grupo aleatorio dependiente se puede obtener de la siguiente forma.

Los elementos impares y pares en cada estrato se separan de la muestra.

$$G_1 = \{\{y_{11}, y_{13}, \ldots, y_{1z_1}\}, \{y_{21}, y_{23}, \ldots, y_{2z_2}\}, \ldots, \{y_{L1}, y_{L3}, \ldots, y_{Lz_L}\}\}$$

$$G_2 = \{\{y_{12}, y_{14}, \ldots, y_{1z_1'}\}, \{y_{22}, y_{24}, \ldots, y_{2z_2'}\}, \ldots, \{y_{L2}, y_{L4}, \ldots, y_{Lz_L'}\}\}$$

si n_h es par $z_h = n_h - 1$ y $z_h' = n_h$, pero si n_h es impar $z_h = n_h$ y $z_h' = n_h - 1$.

Se evalúan los dos estimadores en cada grupo aleatorio, para $\hat{\theta} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} \bar{y}_{hi\cdot}}{n_h}$

$$\hat{\theta}_1 = \hat{\theta}_1(y_{11}, y_{12}, \ldots, y_{1z_1}, y_{21}, y_{22}, \ldots, y_{2z_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lz_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{z_h} \bar{y}_{hi\cdot}}{z_h}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_{11}, y_{12}, \ldots, y_{1z_1'}, y_{21}, y_{22}, \ldots, y_{2z_2'}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lz_L'}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{z_h'} \bar{y}_{hi\cdot}}{z_h'}$$

Si $k=4$ se forman cuatro grupos aleatorios, una de las formas de obtenerlos es separar a los valores de la muestra en 4 grupos.

$$G_1 = \{\{y_{11}, y_{15}, \ldots, y_{1a_1}\}, \{y_{21}, y_{25}, \ldots, y_{2a_2}\}, \ldots, \{y_{L1}, y_{L5}, \ldots, y_{La_L}\}\}$$

$$G_2 = \{\{y_{12}, y_{16}, \ldots, y_{1b_1}\}, \{y_{22}, y_{26}, \ldots, y_{2b_2}\}, \ldots, \{y_{L2}, y_{L6}, \ldots, y_{Lb_L}\}\}$$

$$G_3 = \{\{y_{13}, y_{17}, \ldots, y_{1c_1}\}, \{y_{23}, y_{27}, \ldots, y_{2c_2}\}, \ldots, \{y_{L3}, y_{L7}, \ldots, y_{Lc_L}\}\}$$

$$G_4 = \{\{y_{14}, y_{18}, \ldots, y_{1d_1}\}, \{y_{24}, y_{28}, \ldots, y_{2d_2}\}, \ldots, \{y_{L4}, y_{L8}, \ldots, y_{Ld_L}\}\}$$

y al evaluar el estimador en los cuatro grupos,

$$\hat{\theta}_1 = \hat{\theta}_1(y_{11}, y_{12}, \ldots, y_{1a_1}, y_{21}, y_{22}, \ldots, y_{2a_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{La_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{a_h} \bar{y}_{hi\cdot}}{a_h}$$

$$\hat{\theta}_2 = \hat{\theta}_2(y_{11}, y_{12}, \ldots, y_{1b_1}, y_{21}, y_{22}, \ldots, y_{2b_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lb_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{b_h} \bar{y}_{hi\cdot}}{b_h}$$

$$\hat{\theta}_3 = \hat{\theta}_3(y_{11}, y_{12}, \ldots, y_{1c_1}, y_{21}, y_{22}, \ldots, y_{2c_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Lc_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{c_h} \bar{y}_{hi\cdot}}{c_h}$$

$$\hat{\theta}_4 = \hat{\theta}_4(y_{11}, y_{12}, \ldots, y_{1d_1}, y_{21}, y_{22}, \ldots, y_{2d_2}, \ldots, y_{L1}, y_{L2}, \ldots, y_{Ld_L}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{d_h} \bar{y}_{hi\cdot}}{d_h}$$

con a_h, b_h, c_h, $d_h \le n_h$. Con estos valores se calcula el estimador de varianza, para el otro estimador se procede análogamente.

3.3.4 Muestreo estratificado con arranque aleatorio independiente en cada estrato, cuatro estimadores

El diseño considera a $N_h = k_h.n_h$ para seleccionar uno de los enteros $1, 2, \ldots, k_h$ al azar para cada valor de h. Se genera en forma independiente en cada estrato un arranque aleatorio, la muestra queda integrada por los elementos $i_h, i_h + k_h, i_h + 2k_h, \ldots, i_h + (n_h - 1)k_h$ de cada estrato h.

Los valores de la muestra aplicados a cada estimador se ponderan para obtener cuatro estimadores de la proporción, para generar los grupos aleatorios se procede de manera análoga a un muestreo estratificado.

CAPÍTULO 4. SIMULACIONES (RESULTADOS NUMÉRICOS)

4.1 Introducción

Los programas de cómputo para estimar proporciones y varianzas auxilian de manera importante a la obtención de resultados de manera rápida y confiable. Su diseño requirió de la completa comprensión de los diseños muestrales y de examinar la manera óptima de desarrollarlos con sencillez y claridad.

En este capítulo se analizan los resultados numéricos de simular conteos rápidos. Se comentan los procedimientos de escoger una muestra con y sin reemplazo, así como la manera de realizarlos con diferentes diseños de muestreo. Además, se presentan los resultados numéricos de las simulaciones con diferentes estimadores de la proporción del voto para los estados de Guanajuato, Yucatán, y Michoacán. Se utilizan diferentes tamaños de muestra, criterios de estratificación y las estimaciones de varianza se hacen mediante grupos aleatorios independientes y dependientes.

Se comparan las estimaciones de la proporción del voto con los resultados oficiales, y las estimaciones de varianza usando el método de grupos aleatorios con los errores cuadráticos medios de los estimadores de cada diseño muestral.

Las simulaciones se hicieron con 500 repeticiones, para los estimadores de la proporción se tomaron únicamente aquéllas comprendidas en el rango 0.30 y 0.70, ya que fuera de este intervalo la varianza de la proporción cambia mucho (Cochran, 1977, pag. 53), se necesitarían muestras muy grandes para obtener estimadores precisos del número promedio de cualquier atributo que sea raro en la población. Por ejemplo, el voto a favor de partidos minoritarios.

Primeramente se realizó un estudio piloto para definir el número de simulaciones a realizar, se trabajó con 100, 500, y 1000 para finalmente, al usar el criterio del tiempo de máquina y la diferencia en resultados numéricos, decidir utilizar 500.

En la simulación se utilizó el lenguaje de programación y manejador de bases de datos CLIPPER versión 5.2 (1996) por tener un dominio aceptable de su aplicación. Su compilador incluye rutinas de soporte llamadas 'Microsoft C Floating Point Support Routines'. El hardware empleado tiene las siguientes características : Equipo ACER Procesador 486 , 4 MB en RAM con un tiempo promedio de simulación de 2 minutos para generar estimadores de la proporción y de varianza con grupos aleatorios independientes y dependientes para un sólo diseño y con un sólo conjunto de datos.

4.2 Estructura y descripción de las bases de datos

Las bases de datos que se manejan para simular tienen la misma estructura a la descrita en la sección 3.3 , y se les agrega el siguiente campo.

Campo	Nombre_Campo	Tipo	Rango
1	NUMERO	Numérico	6

NUMERO es el nombre del campo que sirve como auxiliar en la validación de los registros que se escogen al azar de manera que no se repitan.

4.3 Simulación en muestreo aleatorio simple de conglomerados

Para efectuar la simulación se escogió una rutina de tipo recursivo que genera números aleatorios (Bartley, 1987, pag. 331), y se aplica a los procedimientos MUES (P.1, Anexo) y MUEC (P.2, Anexo). MUES significa "MUEstreo Sin Reemplazo" y MUEC "MUEstreo Con Reemplazo " en un MAS, su definición genérica es :

MUEC("nombre_basededatos", N, n), operan de la siguiente manera.

Para generar una m.a.s.s.c.r. se requiere como datos de entrada: el nombre de una base de datos , el número N de registros (casillas o secciones), y el tamaño n de muestra. El procedimiento genera como resultado una base de datos llamada VALIDO.DBF que se forma de dos maneras diferentes :

a) Si el muestreo es sin reemplazo validará cada uno de los números aleatorios generados, y revisa que no se repita alguno hasta obtener el tamaño de muestra solicitado.

b) Si el muestreo es con reemplazo permitirá las repeticiones para completar la muestra.

Se evalúan tres estimadores :

i) $\hat{\bar{Y}} = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$ estimador media de medias, *ii)* $\hat{\bar{Y}}_u = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}}$ estimador insesgado de la proporción poblacional , y *iii)* $\hat{\bar{Y}}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$ estimador de razón

donde $\bar{y}_{i\cdot} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i} = \frac{y_i}{M_i}$ con $y=1,2,..,n$

si el PAN fuera el partido considerado, $y_i = PAN_i$ con $i=1,2,...,N$ y $M_i = VALIDOS_i$.

Con el programa 3 (P 3, Anexo) las estimaciones se hacen simultáneamente para todos los partidos y sus resultados se almacenan en una base de datos. Los resultados se promedian para cada partido para obtener las tablas reportadas.

La base de datos VALIDO.DBF se renombra como MMAS.DBF para utilizarla posteriormente al calcular los grupos aleatorios. El nombre asignado significa " Muestra de un MAS ".

Se debe mencionar que para el cálculo numérico se hicieron algunas simplificaciones para evitar realizar operaciones con un costo relacionado al error numérico de redondeo en el sistema de punto flotante y por facilidad en su manejo.

4.3.1. Simulación para grupos aleatorios independientes con k=2,4

Como la base de datos MMAS.DBF contiene una m.a.s.s.r. se le aplica el programa 4 (P.4, Anexo), mediante el cual se escogen muestras con reemplazo de tamaño $n/2$ (o $n/4$ con el programa 5 (P.5 Anexo)) k veces. Cada una de las k veces el estimador se evalúa y el resultado se guarda en una base de datos, se calcula el promedio de los k valores y el estimador de varianza con las fórmulas ya presentadas. Los resultados se almacenan a una base de datos de acuerdo al estimador y al partido para obtener su promedio.

4.3.2 Simulación para grupos aleatorios dependientes con k=2,4

La base de datos MMAS.DBF se divide en 2 (o 4) bases de datos con el programa 6 (P.6, Anexo) o el 7 (P.7, Anexo), según el valor de k. El estimador se evalúa para obtener 2 (o 4) valores del mismo, que se promedian para calcular el estimador de varianza. El resultado se guarda en una base de datos según el estimador y partido.

4.4 Simulación en un muestreo sistemático de conglomerados

Con el número de registros N y el tamaño de muestra n se aplica el programa 8 (P.8, Anexo) con el que se obtiene $k=N/n$. Se elige un número entero i entre 1 y k utilizando la rutina de generación de una muestra de tamaño 1, con la que se genera VALIDO.DBF con un sólo valor i para que a partir de éste se forme la muestra con los elementos i, $i+k$, $i+2k, \ldots, i+(n-1)k$ y guardarla en VALIDO.DBF sobreescribiéndola. A partir de este momento se le considera una muestra aleatoria simple y se evalúa el estimador correspondiente.

Los resultados numéricos se guardan en una base de datos para obtener su promedio.

Un procedimiento equivalente para simular bajo este diseño es escoger un entero entre 1 y k, esto se hace R veces. Se evalúan los diferentes estimadores $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$, donde $\bar{y}_i$ es la proporción de la muestra cuyo primer elemento es y_i, esto es

$\bar{y}_i = \dfrac{y_i + y_{i+k} + y_{i+2k} + \ldots + y_{i+(n-1)k}}{n}$ para formar $\bar{y}_R = \dfrac{\sum_{i=1}^{k} n_i \bar{y}_i}{R}$ y n_i es el número de veces que se elige el número i de las R veces ($R = \sum_{i=1}^{k} n_i$). De esta manera se realizarían las simulaciones en menos tiempo.

Es posible considerar a $\bar{y}_S = \dfrac{\sum_{i=1}^{k} \bar{y}_i}{k}$ y se observa que aunque no son iguales numéricamente, el valor esperado de $\bar{y}_R$ es $\bar{y}_S$. Para ver esto, considérese que n_i es una variable aleatoria que toma los valores de 1 a k y $E(n_i)=1/k$, entonces

$$E(\bar{y}_R) = E\left(\frac{\sum_{i=1}^{k} n_i \bar{y}_i}{\sum_{i=1}^{k} n_i}\right) = \frac{\sum_{i=1}^{k} \bar{y}_i}{k} = \bar{y}_S .$$

La simulación de grupos aleatorios independientes y dependientes con k=2,4, se realiza de manera análoga al caso descrito para un MAS.

Cuando se tiene una población ordenada, es decir, el orden es importante como el caso del muestreo proporcional al tamaño, el número total de muestras sistemáticas diferentes es k. Una simulación en este diseño se ve limitada a k muestras diferentes equiprobables, esta peculiaridad es fácil de ver si se considera al muestreo sistemático como un caso particular del muestreo por conglomerados, donde los conglomerados son de igual tamaño y se toma una muestra aleatoria de un sólo conglomerado.

4.5. Simulación en un muestreo estratificado de conglomerados

Primeramente se crean un número determinado (L) de bases de datos, se obtienen de dividir la base de datos original en estratos. Por ejemplo, en el caso de estratificar por distrito electoral se generan ED1.DBF, ED2.DBF, ..., y EDL.DBF.

Como se conoce el tamaño de muestra n y se cuenta con los datos N, N_1 , N_2,...,N_L , es posible conocer el valor de la muestra n_k con asignación proporcional en cada estrato.

El programa 9 (P.9, Anexo) obtiene una m.a.s.s.r. en cada estrato y evalúa los estimadores siguientes.

i) $\hat{\bar{Y}}_{st} = \frac{1}{N\bar{M}}\sum_{h=1}^{L} N_h \frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h}$ estimador insesgado, *ii)* $\hat{\bar{Y}}_{mst} = \sum_{k=1}^{L} \frac{N_k}{N} \frac{\sum_{i=1}^{n_k} \bar{y}_{ki\cdot}}{n_k}$ estimador media de medias por estrato, *iii)* $\hat{\bar{Y}}_{ust} = \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h\bar{M}_h}$ estimador insesgado por estrato, y

$\hat{\bar{Y}}_{rst} = \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$ estimador de razón en cada estrato.

Los valores de la media en cada muestra por estrato se asignan a L variables, para ponderarlos, y obtener el estimador de la proporción bajo el diseño. Los resultados se guardan en una base de datos para después promediar las repeticiones.

4.5.1 Simulación para grupos aleatorios independientes con k=2,4

Como se cuenta con una muestra estratificada a partir de ella, con el programa 10 (P.10, Anexo) se escogen m.a.s.s.r de tamaño $n_h/2$ (si k=2) de cada estrato. Si k=4 se escogen 4 muestras de tamaño $n_h/4$ en cada estrato con el programa 11(P.11, Anexo).

Para cada caso (k=2 o 4), se generan k valores del estimador y se calcula el valor promedio para estimar la varianza del estimador, por último se promedian las repeticiones.

4.5.2 . Simulación para grupos aleatorios dependientes con k=2,4

Si k=2 el programa 12 (P. 12, Anexo) se aplica a la muestra en cada estrato, la muestra se divide en 2 grupos (elementos pares e impares) para obtener (después de ponderar cada grupo) el estimador promedio y el estimador de varianza.

Cuando k=4 con las muestras de cada estrato se forman 4 grupos, los valores del estimador para cada grupo se ponderan para obtener el estimador del diseño. Se utiliza el programa 13 (P.13, Anexo) para obtener los estimadores de varianza, y finalmente promediarlos.

4.6. Simulación en un muestreo estratificado y sistemático con arranque aleatorio independiente en cada estrato

Para cada base de datos del estrato correspondiente el programa 14 (P.14, Anexo) calcula k_h-N_h/n_h y genera un número aleatorio i_h , para formar la muestra en ese estrato, evalúa el estimador y obtiene el estimador de muestreo estratificado.

De manera análoga para los grupos aleatorios independientes y dependientes con k=2,4 se procede según lo descrito en 4.5.1 y 4.5.2.

4.7 Resultados

En esta sección se presentan los resultados numéricos obtenidos de las simulaciones de conteos rápidos en los estados de Guanajuato, Yucatán y Michoacán bajo los 3 diferentes diseño de muestreo: MAS, sistemático y estratificado

4.7.1 Estado de Guanajuato

4.7.1.1 Muestreo aleatorio simple de conglomerados de diferente tamaño

El tamaño de muestra para todos los estados y para fines de simulación se decidió con base en los tamaños usados en la unidad de muestreo, el nivel de precisión deseado y los costos de operación en conteos rápidos realizados por el Centro de Estudios de Opinión de la Universidad de Guadalajara.

La división electoral del estado de Guanajuato está constituida por 3005 secciones electorales de las cuales se escoge una muestra de 120 secciones (4 %). Se consideran tres estimadores de la proporción de la votación para los partidos participantes.

Los resultados se presentan en la tabla 2 (pag. 81), se puede observar que el estimador de menor precisión es el estimador media de medias y el estimador que se acerca más al resultado que obtuvieron todos los partidos es el estimador insesgado, aunque el estimador de razón en algunos casos lo supera cuando estima la proporción de votos para el Partido de Acción Nacional y el Partido de la Revolución Democrática.

Los resultados de la estimación de varianza están contenidos en la tabla 8 (pag. 87), se puede advertir que en general, el método de grupos aleatorios para este diseño da una buena estimación de la varianza.

Se puede observar que el método de estimación que mejor se aproxima en la mayoría de los casos es el de grupos aleatorios independientes con $k=4$. Este resultado tendería a confirmar que mientras más grande sea el número k en los grupos aleatorios independientes menor deberá ser el sesgo de su estimador (Wolter, 1985, pag. 34).

4.7.1.2 Muestreo sistemático de conglomerados de diferente tamaño

Se calcula a $k=3005/120=25$ y se elige al azar un número i entre 1 y 25 para luego generar la muestra , los resultados se registran en la tabla 3 (pag. 82) en donde se puede apreciar que los resultados son similares al diseño del MAS es decir, el estimador que menos se aproxima al valor verdadero es el estimador media de medias y los que mejor lo hacen son el estimador insesgado y el estimador consistente (que en algunos casos consigue una mejor aproximación).

La estimación de varianza se reporta en la tabla 9 (pag. 88), los resultados dan buenas estimaciones para la varianza de los estimadores y los grupos aleatorios independientes con $k=2$ son los que mejor la estiman, sin embargo el caso $k=4$ le sigue como el mejor estimador en todos los casos. Para comparar los resultados se utiliza la diferencia absoluta del valor poblacional θ y sus estimador $\hat{\theta}$ por 10^4 es decir $|\theta - \hat{\theta}| x 10^4$.

4.7.1.3 Muestreo estratificado por distrito electoral

El tamaño de muestra es $n=120$, debe observarse que los tamaños de muestra n_i son pequeños y dado que se tienen 18 distritos electorales locales, el tamaño de muestra global es menor o igual que N_i el número de unidades en cada estrato.

Los resultados de las estimaciones, tamaño de muestra y de los 18 estratos, se presentan en la tabla 4 (pag. 83). Son cuatro los estimadores que se evaluaron; el estimador insesgado es el que mejor aproxima al resultado correcto, mientras que el estimador insesgado por estrato consigue superar a los otros dos. El estimador media de medias en cada estrato se aleja más del valor correcto.

Los resultados de estimar la varianza se describen en la tabla 10 (pag. 89), los grupos aleatorios independientes con $k=4$ estiman mejor, aunque los grupos aleatorios dependientes con $k=2$ obtienen la siguiente mejor estimación. El tamaño de muestra por estrato es muy bajo, es decir, n_h es pequeño, al partir la muestra se hacen aún más pequeños que es quizá la razón de que haya diferencias grandes.

4.7.1.4 Muestreo estratificado por distrito electoral y sistemático con arranque aleatorio independiente en cada estrato

Para este diseño de muestreo se utiliza el procedimiento enunciado en la sección de muestreo sistemático (sección 1.5.3.2), se escoge de cada estrato en forma independiente un número entre 1 y k_h, donde h es el estrato h-ésimo de la población y la muestra se genera en cada estrato, para después considerarla como una muestra aleatoria simple, el estimador se pondera en cada estrato para obtener el estimador del diseño.

En la tabla 5 (pag. 84) se reportan los resultados de las simulaciones, tamaños de muestra y de estratos.

El estimador insesgado es el que mejor se aproxima al valor verdadero, aunque no lo es en un diseño estratificado por distrito electoral, el estimador insesgado en cada estrato es el siguiente en aproximación, donde supera al del diseño estratificado por distrito electoral.

El estimador de razón por estrato para este caso en particular no tan sólo supera al estratificado por distrito electoral, para algunos partidos consigue aumentar su precisión con respecto, incluso, al insesgado en cada estrato.

4.7.1.5 Muestreo estratificado por ubicación geográfica urbana/rural de casillas

Los resultados de este diseño se presentan en la tabla 6 (pag. 85), muestran que el estimador con mejor aproximación es el estimador insesgado seguido del estimador insesgado en cada estrato, el estimador de razón por estrato y el estimador media de medias.

La estratificación urbana/rural produce mejores resultados que la estratificación por distrito electoral, ya que supera a prácticamente todos los estimadores.

La estimación de varianza por grupos aleatorios se presenta en la tabla 11 (pag. 90), donde los grupos aleatorios independientes con $k=4$ son los que más se aproximan al verdadero valor del error cuadrado medio.

4.7.1.6 Muestreo estratificado por ubicación geográfica urbana/rural de casillas y sistemático con arranque aleatorio independiente en cada estrato

El procedimiento es análogo al del diseño de muestreo estratificado por distrito electoral con arranque aleatorio independiente en cada estrato, pero la estratificación es por ubicación geográfica urbana / rural de casillas.

Los resultados para este diseño, tamaño de estratos y muestras se detallan en la tabla 7 (pag. 86). Son prácticamente los mismos que los del muestreo estratificado urbana/rural, aunque el estimador insesgado supera al del muestreo estratificado urbano/rural.

Debe mencionarse que salvo los estimadores insesgados para muestreos estratificados, los estimadores no son superiores en general a los correspondientes para los diseños de MAS y sistemático.

4.7.2 Estado de Yucatán

4.7.2.1 Muestreo aleatorio simple de conglomerados de diferente tamaño

Para el diseño se utilizó una población de 1054 secciones electorales, se tomó una m.a.s.s.r. de 95 de ellas (9 %). Los resultados de las simulaciones aparecen en la tabla 12 (pag. 91).

El estimador de razón consiguió mejores resultados seguido por el estimador insesgado y el estimador media de medias.

Los resultados de la estimación de varianza se reportan en la tabla 15 (pag. 94) se observa que en general, los grupos aleatorios independientes con $k=4$ tienen los mejores resultados.

4.7.2.2 Muestreo sistemático de conglomerados de diferente tamaño

Para este diseño $k=N/n=11$ y se procede de manera análoga al del estado de Guanajuato.

Los estimadores insesgado y de razón tiene resultados muy parecidos y consiguen una mejor aproximación que el estimador media de medias.

En la tabla 16 (pag. 95) los resultados de estimar la varianza señalan que los grupos aleatorios independientes con $k=4$ se aproximan mejor a los valores verdaderos, seguidos en precisión por los grupos aleatorios independientes con $k=2$.

4.7.2.3. Muestreo estratificado por distrito electoral

Al considerar $n=95$ (9 %) es posible darse cuenta que los tamaños de muestra n_i son pequeños y el tamaño de muestra global es menor o igual que N_i el número de unidades en cada estrato.

En la tabla 14 (pag. 93) se muestran los resultados, tamaños de los 15 estratos y de las muestras. Se hacen las siguientes observaciones: el estimador insesgado estratificado es de mayor precisión que el estimador insesgado en un MAS, sin embargo los demás estimadores no son más precisos que los del MAS.

4.7.3 Estado de Michoacán

4.7.3.1 Muestreo aleatorio simple de conglomerados de diferente tamaño

Para este estado se considera una población de 2081 secciones electorales (debe recordarse que la base de datos tiene sólo 13 distritos electorales con 2081 secciones, aunque el estado se dividió en 18 distritos).

Los parámetros poblacionales se refieren a la subpoblación de 13 distritos con 2081 secciones, de las que se escoge una m.a.s.s.r. de tamaño 125 (6 %) y se generan los resultados que se presentan en la tabla 18 (pag. 97). Aquí se puede observar que los estimadores insesgado y de razón obtienen más precisión que el estimador media de medias.

4.7.3.2 Muestreo sistemático de conglomerados de diferente tamaño

Para el diseño $k=N/n=16$ y el procedimiento es análogo a los demás estados.

El resultado de las simulaciones se da a conocer en la tabla 19 (pag. 98) que es similar a la del MAS, aunque con menor precisión.

La estimación de varianza se reporta en la tabla 22 (pag. 101), se observa que cuando $k=2$ los grupos aleatorios independientes se acercan mejor al valor verdadero, aunque los resultados son muy parecidos al caso $k=4$.

4.7.3.3 Muestreo estratificado por distrito electoral

Los resultados bajo este diseño se describen en la tabla 20 (pag. 99), se deduce que el estimador insesgado tiene los mejores resultados de estimación; los estimadores insesgado en cada estrato y de razón en cada estrato obtienen resultados parecidos, y en último lugar está el estimador media de medias por estrato.

Excepto por la estimación para el Partido de Acción Nacional se podria decir que el estimador insesgado es tan preciso como el correspondiente al MAS, aunque para los demás partidos no ocurre de esta manera.

En la tabla 23 (pag. 102) se reportan los resultados referentes a la estimación de la varianza para los tres partidos mencionados, se observa que en forma general los resultados de los grupos aleatorios dependientes con k=4 aproximan mejor a los valores verdaderos .

4.8 Cuadro comparativo de conteos rápidos realizados en los tres estados

El siguiente cuadro (Cuadro A) perrmite comparar los resultados de conteos rápidos realizados por Centro de Estudios de Opinión de la Universidad de Guadalajara (CEO) en los estados de Guanajuato, Yucatán y Michoacán con los resultados oficiales.

PARTIDO	GUANAJ UATO		YUCA TÁN		MICHOACÁN	
	Conteo rápido extraoficial	Oficial *	Conteo rápido extraoficial	Oficial **	Conteo rápido extraoficial	Oficial ***
PRI	30.09	32.90	47.91	49.40	39.72	38.80
PAN	60.72	58.10	45.78	43.70	27.20	25.40
PRD	6.77	7.02	3.06	3.10	30.56	32.30
OTRO	1.46	1.98	4.20		2.52	3.18
% Participación	60.42	57.55	69.90	60.03	58.56	60.14(1)
Votos válidos	51900	1244955	47915	503137	60914	886794(1)
Secciones	107	3005	93	1054	129	2081(1)

CUADRO A Cuadro comparativo de conteos rápidos realizados en 3 estados en 1995

* ** periódico " El Nacional ", 28 de mayo de 1995

*** periódico " El Nacional ", 17 de noviembre de 1995

(1) Ver nota aclaratoria, página 53

4.9 Comparación de los métodos muestreo aleatorio simple, muestreo sistemático y muestreo estratificado por distrito electoral.

Para comparar los diseños MAS, sistemático y estratificado por distrito electoral en los estados de Guanajuato, Yucatán y Michoacán se promediaron las diferencias de cada estimador y partido con respecto al valor poblacional. Por ejemplo, para Guanajuato se utiliza la información de la tabla 2 que está entre paréntesis, estos valores se promedian para obtener así una medida de comparación.

Los diseños en los que sus estimadores se aproximan con menos error al valor poblacional son: el muestreo aleatorio simple, el muestreo sistemático y el muestreo estratificado por distrito electoral.

En Guanajuato en particular, el MAS resultó tan preciso como el muestreo sistemático y mejor que el muestreo estratificado por distrito electoral. En Yucatán el MAS fue mejor que el muestreo sistemático y el estratificado por distrito, en Michoacán se observó que el muestreo estratificado por distrito electoral supera ligeramente al muestreo sistemático y al MAS.

4.10 TABLAS

ESTADO DE GUANAJUATO

MUESTREO ALEATORIO SIMPLE DE CONGLOMERADOS DE DIFERENTE TAMAÑO
N=3005 secciones
n=120 (4 %)

500 ensayos

Estimador / Partido	1: $\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$	2: $\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}}$	3: $\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$	Resultados poblacionales
PRI	0.3562, (272)	0.3284, (6)	0.3277, (13)	0.3290
PAN	0.5462, (348)	0.5840, (30)	0.5822, (12)	0.5810
PRD	0.0768, (66)	0.0704, (2)	0.0702, (0)	0.0702
OTROS	0.0207, (9)	0.0199, (1)	0.0199, (1)	0.0198

Tabla 2. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo aleatorio simple comparando con los resultados poblacionales. (1) estimador media de medias, sesgado, (2) estimador insesgado, (3) estimador de razón, sesgado pero consistente
Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE GUANAJUATO

MUESTREO SISTEMÁTICO DE CONGLOMERADOS DE DIFERENTE TAMAÑO N=3005 secciones n=120 (4 %)	1	2	3	500 ensayos
Estimador / Partido	$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$	Resultados poblacionales
PRI	0.3578, (288)	0.3287, (3)	0.3284, (6)	0.3290
PAN	0.5452, (358)	0.5827, (17)	0.5820, (10)	0.5810
PRD	0.0763, (61)	0.0699, (3)	0.0698, (4)	0.0702
OTROS	0.0207, (9)	0.0199, (1)	0.0199, (1)	0.0198

Tabla 3. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo sistemático comparando con los resultados poblacionales. (1) estimador media de medias, sesgado, (2) estimador insesgado, (3) estimador de razón, sesgado pero consistente
Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE GUANAJUATO

MUESTREO ESTRATIFICADO POR DISTRITO ELECTORAL

$N=3005$, $n=118$, $L=18$

$N_1=171$, $N_2=253$, $N_3=160$, $N_4=185$, $N_5=126$, $N_6=177$, $N_7=161$, $N_8=161$, $N_9=135$, $N_{10}=91$, $N_{11}=101$, $N_{12}=212$, $N_{13}=169$, $N_{14}=246$, $N_{15}=161$, $N_{16}=184$, $N_{17}=124$, $N_{18}=188$

$n_1=7$, $n_2=10$, $n_3=6$, $n_4=7$, $n_5=5$, $n_6=7$, $n_7=6$, $n_8=6$, $n_9=5$, $n_{10}=4$, $n_{11}=4$, $n_{12}=8$, $n_{13}=7$, $n_{14}=10$, $n_{15}=6$, $n_{16}=7$, $n_{17}=5$, $n_{18}=8$

500 ensayos

Estimador / Partido	1: $\sum_{h=1}^{L}\frac{N_h}{N\bar{M}}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h}$	2: $\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} \bar{y}_{hi\cdot}}{n_h}$	3: $\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h\bar{M}_h}$	4: $\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$	Resultados poblacionales
PRI	0.3286, (4)	0.3564, (274)	0.3378, (88)	0.3394, (104)	0.3290
PAN	0.5846, (36)	0.5466, (344)	0.5711, (99)	0.5663, (147)	0.5810
PRD	0.0700, (2)	0.0764, (62)	0.0739, (37)	0.0742, (40)	0.0702
OTROS	0.0198, (0)	0.0206, (8)	0.0201, (3)	0.0201, (3)	0.0198

Tabla 4. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito electoral comparando con los resultados poblacionales. (1) estimador insesgado, (2) estimador media de medias por estrato, sesgado, (3) estimador insesgado por estrato, sesgado, (4) estimador de razón por estrato, sesgado

Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE GUANAJUATO

MUESTREO ESTRATIFICADO POR DISTRITO ELECTORAL Y SISTEMÁTICO CON ARRANQUE ALEATORIO INDEPENDIENTE EN CADA ESTRATO N=3005, n=118, L=18 $N_1=171$, $N_2=253$, $N_3=160$, $N_4=185$, $N_5=126$, $N_6=177$, $N_7=161$, $N_8=161$, $N_9=135$, $N_{10}=91$, $N_{11}=101$, $N_{12}=212$, $N_{13}=169$, $N_{14}=246$, $N_{15}=161$, $N_{16}=184$, $N_{17}=124$, $N_{18}=188$ $n_1=7$, $n_2=10$, $n_3=6$, $n_4=7$, $n_5=5$, $n_6=7$, $n_7=6$, $n_8=6$, $n_9=5$, $n_{10}=4$, $n_{11}=4$, $n_{12}=8$, $n_{13}=7$, $n_{14}=10$, $n_{15}=6$, $n_{16}=7$, $n_{17}=5$, $n_{18}=8$					
	1	2	3	4	500 ensayos
Estimador / Partido	$\sum_{h=1}^{L}\frac{N_h}{N\overline{M}}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{n_h\overline{M}_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h}M_{hi}}$	Resultados poblacionales
PRI	0.3277, (13)	0.3545, (255)	0.3370, (80)	0.3363, 73	0.3290
PAN	0.5896, (86)	0.5495, (315)	0.5760, (50)	0.5706, 104	0.5810
PRD	0.0697, (5)	0.0754, (52)	0.0738, (36)	0.0731, 29	0.0702
OTROS	0.0198, (0)	0.0206, (8)	0.0201, (3)	0.0200, 2	0.0198

Tabla 5. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito electoral y sistemático con arranque aleatorio independiente en cada estrato, comparando con los resultados poblacionales. (1) estimador insesgado, (2) estimador media de medias por estrato, sesgado, (3) estimador insesgado por estrato, sesgado, (4) estimador de razón por estrato, sesgado

Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

OTROS	(0)	0.020	0.0200

Tabla 5. Resulta[illegible] ensayos de sitio [illegible] un diseño [illegible] estratificado po[illegible] sistemático con [illegible] aleatorio independiente [illegible] comparando con [illegible] s poblacionale[illegible] esperados [illegible] de medias [illegible] estimador [illegible]

[illegible] diferencias [illegible] poblaciona[illegible]

ESTADO DE GUANAJUATO

MUESTREO ESTRATIFICADO URBANO / RURAL DE CASILLAS

$N=4399$, $n=176$
$N_1=2856$, $N_2=1543$
$n_1=114$, $n_2=62$

500 ensayos

Estimador / Partido	1: $\sum_{h=1}^{L}\frac{N_h}{N\overline{M}}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{n_h}$	2: $\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}$	3: $\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{n_h\overline{M}_h}$	4: $\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h}M_{hi}}$	Resultados poblacionales
PRI	0.3287, (3)	0.3438, (148)	0.3383, (93)	0.3374, (84)	0.3290
PAN	0.5835, (25)	0.5596, (214)	0.5716, (94)	0.5700, (110)	0.5810
PRD	0.0707, (5)	0.0762, (60)	0.0727, (25)	0.0725, (23)	0.0702
OTROS	0.0198, (0)	0.0203, (5)	0.0201, (3)	0.0200, (2)	0.0198

Tabla 6. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por ubicación geográfica de casillas urbana / rural comparando con los resultados poblacionales. (1) estimador insesgado, (2) estimador media de medias por estrato, sesgado, (3) estimador insesgado por estrato, sesgado, (4) estimador de razón por estrato, sesgado Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE GUANAJUATO

MUESTREO ESTRATIFICADO URBANO/ RURAL DE CASILLAS Y SISTEMÁTICO CON ARRANQUE ALEATORIO INDEPENDIENTE EN CADA ESTRATO

N=4399, n=176
N_1=2856, N_2=1543
n_1=114, n_2=62

	1	2	3	4	500 ensayos
Estimador / Partido	$\sum_{h=1}^{L}\frac{N_h}{N\overline{M}}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h\overline{M}_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$	Resultados poblacionales
PRI	0.3294, (4)	0.3456, (166)	0.3391, (101)	0.3387, (97)	0.3290
PAN	0.5823, (13)	0.5582, (228)	0.5703, (107)	0.5692, (118)	0.5810
PRD	0.0703, (1)	0.0759, (57)	0.0722, (20)	0.0721, (19)	0.0702
OTROS	0.0198, (0)	0.0203, (5)	0.0200, (2)	0.0200, (2)	0.0198

Tabla 7. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por ubicación geográfica de casillas urbana/ rural con arranque aleatorio independiente en cada estrato comparando con los resultados poblacionales (1) estimador insesgado, (2) estimador media de medias por estrato, sesgado, (3) estimador insesgado por estrato, sesgado, (4) estimador de razón por estrato, sesgado
Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE GUANAJUATO

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN, CALCULADA PARA EL MÉTODO DE GRUPOS ALEATORIOS MUESTREO ALEATORIO SIMPLE DE CONGLOMERADOS CON DIFERENTE TAMAÑO N=3005 secciones n=120 (4%) 1 2 — 500 ensayos					
Estimador	$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$, $\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$				
	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	2.35459, 1.78932	2.27385, 1.75459	2.25019, 1.70250	2.44485, 1.82542	2.19725, 1.65485
PAN	2.71872, 2.25882	2.66356, 2.29522	2.66416, 2.21645	2.87815, 2.27647	2.54569, 2.09986

Tabla 8. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo aleatorio simple para dos estimadores, comparando con las fórmulas del error cuadrático medio. (1) estimador media de medias, sesgado, (2) estimador de razón, sesgado pero consistente

ESTADO DE GUANAJUATO

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN, CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS MUESTREO SISTEMÁTICO DE CONGLOMERADOS DE DIFERENTE TAMAÑO N=3005 secciones n=120 (4%) 1 2 500 ensayos					
Estimador X 10^4	$\frac{\sum_{i=1}^{n} \overline{y}_{i\cdot}}{n}$, $\frac{\sum_{i=1}^{n} M_i \overline{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$				
	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	2.20283,1.77595	3.50623, 1.51492	2.25854, 1.78913	3.22221, 2.02945	2.19725,1.65485
PAN	2.55653, 2.22575	2.25926, 1.12009	2.64960, 2.27137	2.91870, 1.81630	2.54569,2.09986

Tabla 9. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo sistemático para dos estimadores, comparando con las fórmulas de E.C.M*. (1) estimador media de medias, sesgado, (2) estimador de razón sesgado pero consistente. * Error cuadrático medio para un muestreo aleatorio simple

ESTADO DE GUANAJUATO

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN, CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS
MUESTREO ESTRATIFICADO POR DISTRITO DE CONGLOMERADOS DE DIFERENTE TAMAÑO
N=3005 secciones
n=120 (4%)

1 2 — 500 ensayos

$$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}, \quad \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h}M_{hi}}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	1.44611, 1.41456	1.97455, 1.78015	1.75839, 1.74383	2.17368, 2.14970	1.87331, 1.62753
PAN	1.52511, 1.52686	2.17615, 2.04467	1.87563, 1.89056	2.36217, 2.37083	2.02453, 1.88770

Tabla 10. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito para dos estimadores, comparando con las fórmulas de E.C.M*. (1) estimador media de medias por estrato, sesgado, (2) estimador de razón por estrato, sesgado. *Error cuadrático medio

ESTADO DE GUANAJUATO

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN, CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS
MUESTREO ESTRATIFICADO URBANO / RURAL DE CASILLAS
N=4399 casillas
n=176 (4%)

1 2 500 ensayos

$$\sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} \overline{y}_{hi\cdot}}{n_h}, \quad \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} M_{hi}\overline{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	1.04069, 1.03436	1.06509, 1.10692	0.99998, 0.97382	1.04311, 1.06638	1.02024, 1.01871
PAN	1.19297, 1.15668	1.15129, 1.15781	1.09647, 1.08441	1.13583, 1.14586	1.10342, 1.11403

Tabla 11. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por ubicación geográfica de casillas urbana/ rural para dos estimadores, comparando con las fórmulas de E.C.M*. (1) estimador media de medias por estrato, sesgado, (2) estimador de razón por estrato, sesgado. * Error cuadrático medio

ESTADO DE YUCATÁN

MUESTREO ALEATORIO SIMPLE DE CONGLOMERADOS DE DIFERENTE TAMAÑO N=1054 secciones n=95 (9 %)	1	2	3	500 ensayos
Estimador / Partido	$\frac{\sum_{i=1}^{n} \bar{y}_i \cdot}{n}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_i \cdot}{n\bar{M}}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_i \cdot}{\sum_{i=1}^{n} M_i}$	Resultados poblacionales
PRI	0.5201, (190)	0.5036, (25)	0.4996, (15)	0.5011
PAN	0.4355, (198)	0.4598, (45)	0.4560, (7)	0.4553
PRD	0.0334, (3)	0.0346, (9)	0.0342, (5)	0.0337
OTROS	0.0109, (9)	0.0103 (3)	0.0102, (2)	0.0100

Tabla 12. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo aleatorio simple comparando con los resultados poblacionales. (1) estimador media de medias, sesgado, (2) estimador insésgado, (3) estimador de razón, sesgado pero consistente
Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE YUCATÁN

MUESTREO SISTEMÁTICO DE CONGLOMERADOS DE DIFERENTE TAMAÑO N=1054 secciones n=95 (9 %)	1	2	3	500 ensayos
Estimador / Partido	$\frac{\sum_{i=1}^{n} \bar{y}_{i}.}{n}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i}.}{n\bar{M}}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i}.}{\sum_{i=1}^{n} M_i}$	Resultados poblacionales
PRI	0.5179, (168)	0.4964, (47)	0.4963, (48)	0.5011
PAN	0.4389, (164)	0.4602, (49)	0.4605, (52)	0.4553
PRD	0.0321, (16)	0.0329, (8)	0.0330, (7)	0.0337
OTROS	0.0111, (11)	0.0102, (2)	0.0102, (2)	0.0100

Tabla 13. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo sistemático comparando con los resultados poblacionales. (1) estimador media de medias, sesgado, (2) estimador insesgado, (3) estimador de razón, sesgado pero consistente

Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE YUCATÁN

MUESTREO ESTRATIFICADO POR DISTRITO ELECTORAL

$N=1054$, $n=94$, $L=15$

$N_1=56$, $N_2=62$, $N_3=66$, $N_4=71$, $N_5=59$, $N_6=65$, $N_7=60$, $N_8=69$, $N_9=78$, $N_{10}=81$, $N_{11}=85$, $N_{12}=83$, $N_{13}=68$, $N_{14}=74$, $N_{15}=77$

$n_1=5$, $n_2=6$, $n_3=6$, $n_4=6$, $n_5=5$, $n_6=6$, $n_7=5$, $n_8=6$, $n_9=7$, $n_{10}=7$, $n_{11}=8$, $n_{12}=7$, $n_{13}=6$, $n_{14}=7$, $n_{15}=7$

	1	2	3	4	500 ensayos
Estimador / Partido	$\sum_{h=1}^{L}\frac{N_h}{N\bar{M}}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h\bar{M}_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$	Resultados poblacionales
PRI	0.5075, (64)	0.5373, (362)	0.5260, (249)	0.5265, (254)	0.5011
PAN	0.4578, (25)	0.4417, (136)	0.4561, (8)	0.4531, (22)	0.4553
PRD	0.0371, (34)	0.0374, (37)	0.0384, (47)	0.0382, (45)	0.0337
OTROS	0.0151, (51)	0.0184, (84)	0.0171, (71)	0.0170, (70)	0.0100

Tabla 14. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito electoral comparando con los resultados poblacionales. (1) estimador insesgado, (2) estimador media de medias por estrato, sesgado, (3) estimador insesgado por estrato, sesgado, (4) estimador de razón por estrato, sesgado

Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE YUCATÁN

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN , CALCULADA PARA EL MÉTODO DE GRUPOS ALEATORIOS
MUESTREO ALEATORIO SIMPLE DE CONGLOMERADOS DE DIFERENTE TAMAÑO
N=1054 secciones
n=95 (9%)

1 2 500 ensayos

$$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}, \quad \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	2.50670, 2.45420	2.47427, 2.30008	2.65658, 2.42597	2.66030, 2.41640	2.63828, 2.38371
PAN	2.93287, 2.99617	2..93936 2.96159	3.13248, 3.19425	3.10217, 3.23682	3.01739, 3.11957

Tabla 15. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo aleatorio simple para dos estimadores, comparando con las fórmulas de E.C.M.. (1) estimador media de medias, sesgado, (2) estimador de razón, sesgado pero consistente

ESTADO DE YUCATÁN

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN , CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS
MUESTREO SISTEMÁTICO DE CONGLOMERADOS DE DIFERENTE TAMAÑO
N=1054 secciones
n=95 (9%)

1 2 — 500 ensayos

$$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}, \quad \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios Dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios Dependientes k=4	Error cuadrático medio * por fórmula
PRI	2.51022, 2.44075	3.85275, 3.08670	2.71786, 2.42555	3.58767, 2.78273	2.63828, 2.38371
PAN	2.89226, 3.12209	1.52444, 1.29506	3.22579, 3.25157	3.35049, 2.22647	3.01739, 3.11957

Tabla 16. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo sistemático para dos estimadores, comparando con las fórmulas de varianza. (1) estimador media de medias, sesgado, (2) estimador de razón, sesgado pero consistente
* Error cuadrático medio de un muestreo aleatorio simple

ESTADO DE YUCATÁN

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN , CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS
MUESTREO ESTRATIFICADO POR DISTRITO DE CONGLOMERADOS DE DIFERENTE TAMAÑO
N=1054 secciones
n=95 (4%)

1 2 500 ensayos

$$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\overline{y}_{hi\cdot}}{n_h}\ ,\ \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\overline{y}_{hi\cdot}}{\sum_{i=1}^{n_h}M_{hi}}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	1.44141, 1.40526	1.76582, 1.78121	1.48340, 1.47061	1.86478, 1.89670	1.61672, 1.12600
PAN	1.50528, 1.54900	1.84167, 1.90084	1.41106, 1.45300	2.02032, 2.03410	1.61634, 1.25568

Tabla 17. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito electoral para dos estimadores, comparando con las fórmulas de E.C.M. (1) estimador media de medias por estrato, sesgado, (2) estimador de razón por estrato, sesgado

ESTADO DE MICHOACÁN

MUESTREO ALEATORIO SIMPLE DE CONGLOMERADOS DE DIFERENTE TAMAÑO N=2081 secciones n=125 (6 %)				
	1	2	3	500 ensayos
Estimador / Partido	$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$	Resultados poblacionales
PRI	0.3885, (55)	0.3818, (12)	0.3818, (12)	0.3830
PAN	0.2503, (257)	0.2768, (8)	0.2766, (6)	0.2760
PRD	0.3289, (216)	0.3088, (15)	0.3089, (16)	0.3073
OTROS	0.0322, (15)	0.0328, (9)	0.0328, (9)	0.0337

Tabla 18. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo aleatorio simple, comparando con los resultados poblacionales. (1) estimador media de medias, sesgado, (2) estimador insesgado, (3) estimador de razón, sesgado pero consistente
Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE MICHOACÁN

MUESTREO SISTEMÁTICO DE CONGLOMERADOS DE DIFERENTE TAMAÑO

N=2081 secciones
n=125 (6 %)

	1	2	3	500 ensayos
Estimador / Partido	$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{n\bar{M}}$	$\frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$	Resultados poblaciona les
PRI	0.3890, (60)	0.3813, (17)	0.3816, (14)	0.3830
PAN	0.2518, (242)	0.2780, (20)	0.2783, (23)	0.2760
PRD	0.3250, (177)	0.3054, (19)	0.3055, (18)	0.3073
OTROS	0.0334, (3)	0.0346, (9)	0.0347, (10)	0.0337

Tabla 19. Resultado de 500 ensayos de simulación de conteo rápido, en un diseño de muestreo sistemático, comparando con los resultados poblacionales. (1) estimador media de medias, sesgado, (2) estimador insesgado, (3) estimador de razón, sesgado pero consistente
Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE MICHOACÁN

MUESTREO ESTRATIFICADO POR DISTRITO ELECTORAL $N=2081$, $n=125$, $L=13$ $N_1=229$, $N_2=211$, $N_3=139$, $N_4=145$, $N_5=161$, $N_6=199$, $N_7=138$, $N_8=139$, $N_9=119$, $N_{10}=172$, $N_{11}=128$, $N_{12}=128$, $N_{13}=173$ $n_1=14$, $n_2=13$, $n_3=8$, $n_4=9$, $n_5=10$, $n_6=12$, $n_7=8$, $n_8=8$, $n_9=7$, $n_{10}=10$, $n_{11}=8$, $n_{12}=8$, $n_{13}=10$					
	1	2	3	4	500 ensayos
Estimador / Partido	$\sum_{h=1}^{L}\frac{N_h}{N\overline{M}}\frac{\sum_{i=1}^{n_h} M_{hi}\overline{y}_{hi}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\overline{y}_{hi\cdot}}{n_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\overline{y}_{hi\cdot}}{n_h\overline{M}_h}$	$\sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h} M_{hi}\overline{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$	Resultados poblacionales
PRI	0.3842, (12)	0.3907, (77)	0.3892, (62)	0.3877, (47)	0.3830
PAN	0.2798, (38)	0.2528, (232)	0.2725, (35)	0.2702, (58)	0.2760
PRD	0.3057, (16)	0.3231, (158)	0.3093, (20)	0.3094, (21)	0.3073
OTROS	0.0337, (0)	0.0332, (5)	0.0327, (10)	0.0327, (10)	0.0337

Tabla 20. Resultado de 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito electoral comparando con los resultados poblacionales. (1) estimador insesgado, (2) estimador media de medias por estrato, sesgado, (3) estimador insesgado por estrato, sesgado, (4) estimador de razón por estrato, sesgado. Entre paréntesis aparecen las diferencias entre las estimaciones y el valor poblacional por 10^4

ESTADO DE MICHOACÁN

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN ,CALCULADA PARA EL MÉTODO DE GRUPOS ALEATORIOS

MUESTREO ALEATORIO SIMPLE DE CONGLOMERADOS DE DIFERENTE TAMAÑO

N=2081 secciones

n=125 (6 %)

1 2 500 ensayos

$$\frac{\sum_{i=1}^{n} \overline{y}_{i\cdot}}{n} , \quad \frac{\sum_{i=1}^{n} M_i \overline{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	1.92904, 1.82892	1.77628, 1.67406	1.88340, 1.73115	1.84656, 1.64297	1.77774, 1.61114
PAN	2.77386, 3.38054	2.76906, 3.38807	2.80691, 3.49679	2.80052, 3.40138	2.63680, 3.31918
PRD	2.65030, 2.50218	2.84222, 2.92550	3.00270, 2.95187	2.95606, 2.83506	2.86940, 2.77092

Tabla 21. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo aleatorio simple para dos estimadores, comparando con las fórmulas de E.C.M.* . (1) estimador media de medias, sesgado, (2) estimador de razón, sesgado pero consistente. * Error cuadrático medio

ESTADO DE MICHOACÁN

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN , CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS
MUESTREO SISTEMÁTICO DE CONGLOMERADOS DE DIFERENTE TAMAÑO
N=2081 secciones
n=125 (6 %)

1 2 500 ensayos

$$\frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}, \quad \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	1.95938, 1.67730	1.38488, 0.86940	1.90698, 1.76184	1.31409, 1.11289	1.77774, 1.61114
PAN	2.61900, 3.22701	0.57977, 1.93887	3.10122, 3.88947	0.86932, 2.08830	2.63680, 3.31918
PRD	2.87782, 2.81533	0.88818, 1.96985	2.99326, 2.96657	1.60416, 2.28214	2.86940, 2.77092

Tabla 22. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo sistemático para dos estimadores, comparando con las fórmulas de varianza. (1) estimador media de medias, sesgado, (2) estimador de razón, sesgado pero consistente .* Error cuadrático medio de un muestreo aleatorio simple

ESTADO DE MICHOACÁN

ESTIMACIÓN DE LA VARIANZA PARA DOS DE LOS ESTIMADORES DE LA PROPORCIÓN , CALCULADA POR EL MÉTODO DE GRUPOS ALEATORIOS

MUESTREO ESTRATIFICADO POR DISTRITO DE CONGLOMERADOS DE DIFERENTE TAMAÑO

N=2081 secciones

n=125 (6%)

1 2 500 ensayos

$$\sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} \bar{y}_{hi\cdot}}{n_h} \qquad \sum_{h=1}^{L} \frac{N_h}{N} \frac{\sum_{i=1}^{n_h} M_{hi} \bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}}$$

Estimador X 10^4	Grupos aleatorios independientes k=2	Grupos aleatorios dependientes k=2	Grupos aleatorios independientes k=4	Grupos aleatorios dependientes k=4	Error cuadrático medio por fórmula
PRI	1.16206, 0.99618	1.65313, 1.37756	0.97049, 0.856255	1.66532, 1.50142	1.49610, 1.25338
PAN	1.61873, 1.88405	2.52349, 3.12122	1.40800, 1.58687	2.30269, 2.71107	2.02308, 2.60713
PRD	1.91127, 2.00429	2.92509, 2.99719	1.84135, 1.82261	2.75865, 2.80792	2.58027, 2.62131

Tabla 23. Resultado de la estimación de la varianza en 500 ensayos de simulación de conteo rápido en un diseño de muestreo estratificado por distrito electoral para dos estimadores, comparando con las fórmulas de E.C.M..

(1) estimador media de medias por estrato, sesgado, (2) estimador de razón por estrato, sesgado

CAPITULO 5. CORRELACIÓN LINEAL ENTRE LISTA NOMINAL Y VOTOS VÁLIDOS

5.1 Introducción

En este capitulo se presenta un análisis sobre la correlación entre el número de electores registrados en lista nominal y los votos válidos emitidos para los tres estados de la República.

Se calcularon los coeficientes de la recta de regresión por mínimos cuadrados, la variable explicativa X es la cardinalidad de la lista nominal en la sección y la variable respuesta Y el es número de votos válidos en la sección.

Si $Y=a_0+a_1X$ es la recta de mínimos cuadros,

$$a_0 = \frac{\sum Y \sum X^2 - \sum X \sum XY}{N\sum X^2 - (\sum X)^2}$$

$$a_1 = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

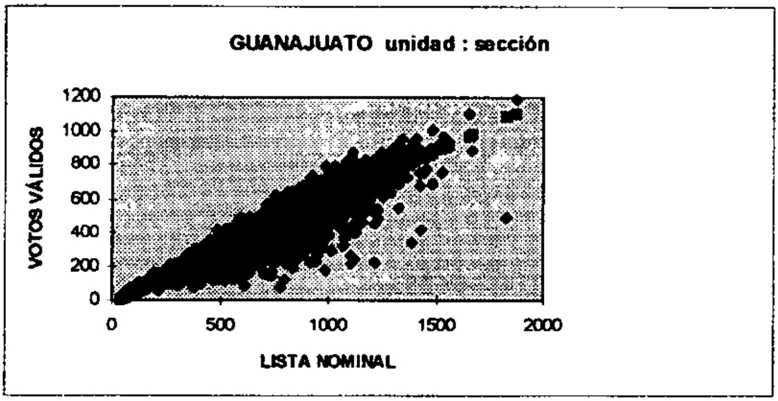
5.2 Caso Guanajuato

Para este estado se podrá comparar la distribución de las listas nominales con los votos válidos emitidos en su totalidad y por distribución geográfica urbana/rural.

5.2.1 Considerando el total de secciones

Se tomaron las 3005 secciones registradas en el estado, para generar el análisis de regresión del que se presenta el resumen, así como la gráfica correspondiente a los datos y su pronóstico.

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.91070127
Coeficiente de determinación R^2	0.829376804
R^2 ajustado	0.829319986
Error típico	77.48656823
Observaciones	3005



Gráfica A. Recta de mínimos cuadrados para la cardinalidad X de la lista nominal en la sección, y Y el número de votos válidos en la sección.

la fórmula de la recta de regresión es

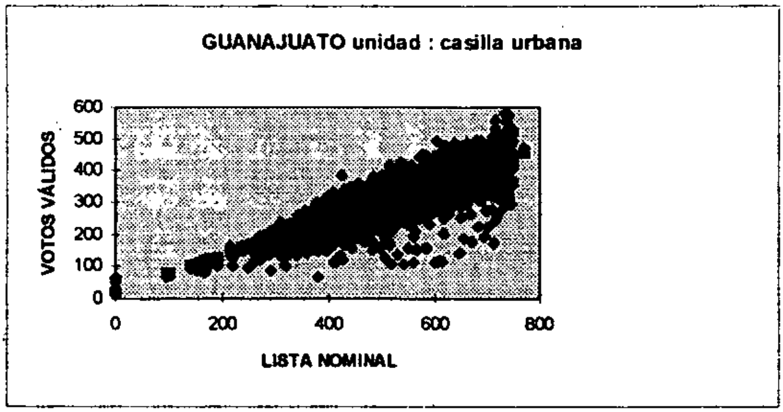
$y = 0.6029x - 19.6751$

0.6029 es la pendiente de la recta, representa el cambio del número de votos válidos dividido por el correspondiente cambio de la cardinalidad de la lista nominal en la sección. Este valor es parecido al porcentaje de participación 0.5755, -19.6751 es el valor de la recta cuando $x=0$

5.2.2 Considerando sólo casillas urbanas

Las casillas son un total de 2856 , se reporta el análisis estadístico, su gráfica correspondiente y fórmula de regresión.

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.846825626
Coeficiente de determinación R^2	0.717113641
R^2 ajustado	0.717014521
Error típico	43.42327144
Observaciones	2856



Gráfica B. Recta de mínimos cuadrados para la cardinalidad X de la lista nominal en la casilla, y Y el número de votos válidos en la casilla Se considera sólo casillas urbanas

la fórmula de regresión es

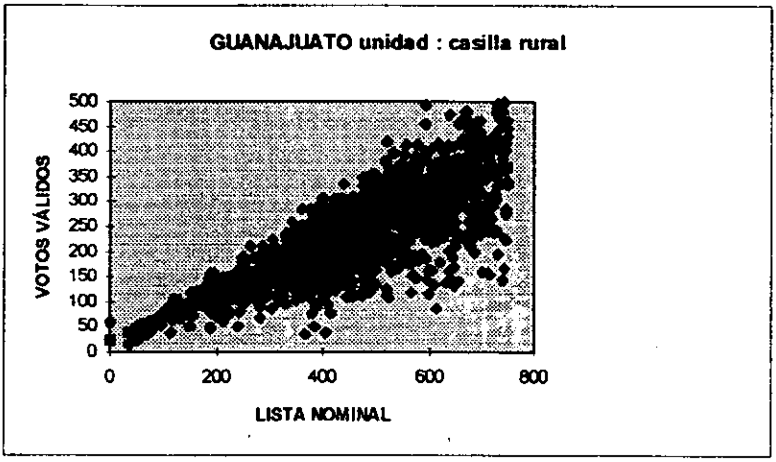
$y = 0.5665x + 20.2839$

0.5665 es la pendiente de la recta, representa el cambio del número de votos válidos dividido por el correspondiente cambio de la cardinalidad de la lista nominal en la sección urbana. Este valor es parecido al porcentaje de participación 0.6063, 20.2839 es el valor de la recta cuando x=0

5.2.3 Considerando sólo casillas rurales

Las casillas rurales son 1543, se genera el siguiente resumen estadístico, gráfica y fórmula de regresión

Estadísticas de la regresión	
Coeficiente de correlación Múltiple	0.781326684
Coeficiente de determinación R^2	0.610471386
R^2 ajustado	0.61021861
Error típico	55.73221266
Observaciones	1543



Gráfica C. Recta de mínimos cuadrados para la cardinalidad X de la lista nominal en la casilla, y Y el número de votos válidos en la casilla. Se considera sólo casillas rurales

la fórmula de regresión es

$$y = 0.4631x - 22.6372$$

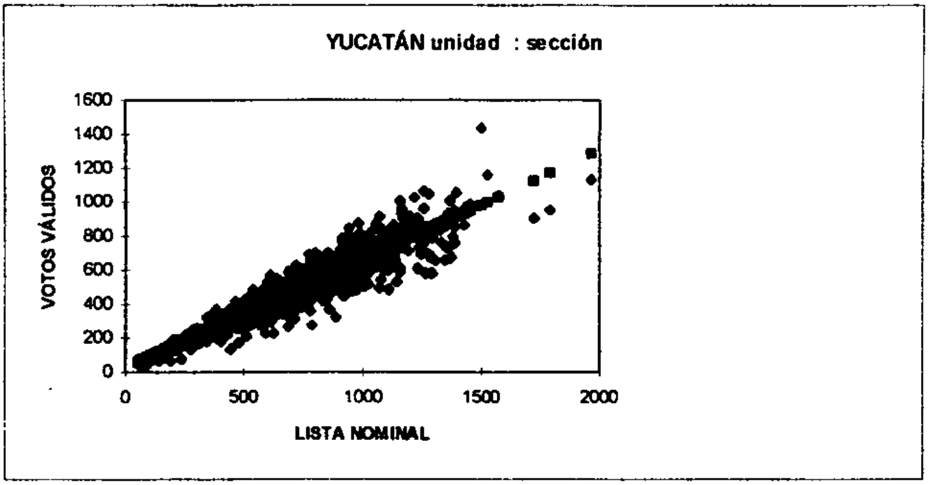
0.4631 es la pendiente de la recta, representa el cambio del número de votos válidos dividido por el correspondiente cambio de la cardinalidad de la lista nominal en la sección rural. Este valor es similar al porcentaje de participación 0.5124, -22.6372 es el valor de la recta cuando x=0.

El porcentaje de participación urbana en Guanajuato fue 60.63 %, el rural 51.24% y el del total 57.55%.

5.3 Caso Yucatán

Para este estado se consideran 1054 secciones, lo que genera el siguiente resumen, gráfica y fórmula de regresión

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.93537069
Coeficiente de determinación R^2	0.874918328
R^2 ajustado	0.874799429
Error típico	72.0887764
Observaciones	1054



Gráfica D. Recta de mínimos cuadrados para la cardinalidad X de la lista nominal en la sección, y Y el número de votos válidos en la sección.

Su fórmula de regresión es :

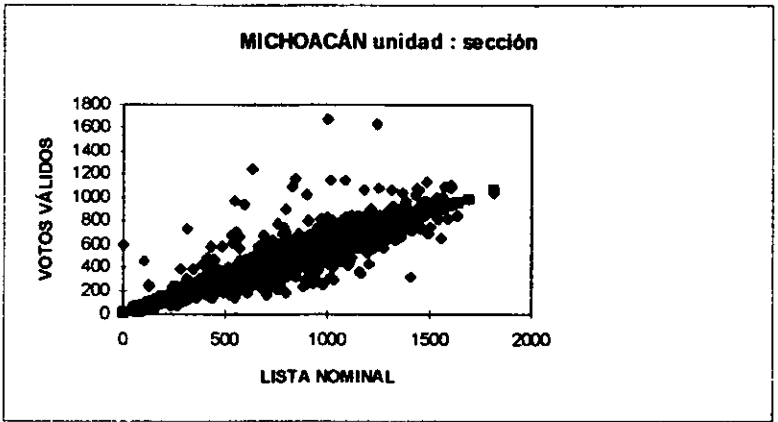
$y = 0.6404x + 22.6985$

0.6404 es la pendiente de la recta, representa el cambio del número de votos válidos dividido por el correspondiente cambio de la cardinalidad de la lista nominal en la sección. Este valor es muy parecido al porcentaje de participación, 22.6985 es el valor de la recta cuando x=0

5.4 Caso Michoacán

Estado con 2081 secciones electorales, su análisis se reporta en la siguiente tabla así como la fórmula de regresión

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.8852209
Coeficiente de determinación R^2	0.78361604
R^2 ajustado	0.78351196
Error típico	0
Observaciones	2081



Gráfica E. Recta de mínimos cuadrados para la cardinalidad X de la lista nominal en la sección, y Y el número de votos válidos en la sección.

y la fórmula de regresión es

$y = 0.5774x + 16.9919$

0.5774 es la pendiente de la recta, representa el cambio del número de votos válidos dividido por el correspondiente cambio de la cardinalidad de la lista nominal en la sección. Este valor es similar al porcentaje de participación 0.6014 , 16.9919 es el valor de la recta cuando x=0.

5.5 Porcentajes de participación

El porcentaje de participación se calculó al dividir el número de votos válidos (los emitidos para alguno de los partidos registrados) y el número de personas registradas en lista nominal, se obtienen los siguientes porcentajes : Guanajuato, 57.55 ; Yucatán, 60.03 y para el estado de Michoacán la registrada en la base de datos fue del 60.14.

Los porcentajes de participación desglosados por estrato son los siguientes: para Guanajuato con estratificación urbana/rural 60.63 % y 51.24 % respectivamente; con estratificación por distrito electoral con 18 estratos 56.71 %, 65.28 %, 63.65 %, 66.71 %, 62.33 %, 48.28 %, 57.97 %, 58.69 %, 60.72 %, 61.73 %, 48.17 %, 48.91 %, 47.11 %, 54.68 %, 60.31 %, 56.51 %, 53.41 %, 55.44 % respectivamente. Para Yucatán con estratificación por distrito electoral y con 15 estratos los porcentajes de participación son: 68.10 %, 66.02 %, 63.29 %, 66.69 %, 62.01 %, 61.20 %, 61.37 %, 60.26 %, 71.60 %, 69.97 %, 70.39 %, 72.25 %, 71.56 %, 68.19 %, 70.75 % respectivamente. En Michoacán se estratificó por distrito electoral en 13 estratos, los porcentajes de participación por estrato son: 58.42 %, 62.87 %, 60.18 %, 61.58 %, 60.73 %, 59.88 %, 59.17 %, 61.13 %, 53.07 %, 58.03 %, 58.45 %, 58.64 %, 67.67 % respectivamente.

Debe mencionarse que existen otras maneras de calcular el porcentaje de participación, por ejemplo, si se toma como votos válidos a todos los votos emitidos incluyendo a los anulados y por candidatos no registrados.

5.6 Comentarios

Para el caso de Guanajuato la participación de los ciudadanos fue concurrida en un porcentaje del 50 % alrededor de la secciones que tenían una lista nominal de entre 500 y 1200 electores, que es donde se concentra el promedio de lista nominal que es de 720. Los datos tienen una tendencia lineal positiva, aunque existen casos aislados de una participación baja con respecto al promedio de 414 de votos válidos.

En la regresión obtenida cuando se considera únicamente a las casillas urbanas se observa que su tendencia también es lineal. Tiene pendiente e intersección similares al caso de casillas rurales, aunque ambas tienden a acumular su participación en las secciones con más cantidad de electores en lista nominal.

La participación es más alta en las casillas rurales con una media de votos válidos de 309, así como un porcentaje de participación más alto con 60.63.

Cuando se considera a todas las casillas la tendencia de la votación válida tiende a reunirse en las secciones con menos número de lista nominal.

Para el estado de Yucatán la participación también registró una votación de tendencia lineal positiva, aunque tiende a no concentrarse uniformemente. Se distribuye a lo largo de la recta y se registra en todos los casos un nivel de participación superior al 50 %, con un promedio de votos válidos de 477.

En Michoacán la participación se distribuye a lo largo de la recta, aunque en algunas secciones su participación es alta, se registra un nivel de participación superior al 50 % y al observado en Yucatán.

CAPÍTULO 6. CONCLUSIONES

La aplicación de los diversos diseños de muestreo se lleva a cabo al considerar como unidad de muestreo al conglomerado de electores llamado sección.

Aunque fueron probados tres diseños de muestreo con algunas variantes, únicamente se considera la comparación de los utilizados en los tres estados como el muestreo aleatorio simple, muestreo sistemático y el muestreo estratificado por distrito electoral.

Para obtener el mejor estimador se promediaron las diferencias absolutas de todos los estimadores con el valor poblacional sin importar el diseño, se observó lo siguiente:

El mejor estimador para los tres estados es el estimador de razón; con un muestreo sistemático para el estado de Guanajuato y bajo un muestreo aleatorio simple en los estados de Yucatán y Michoacán. Resultados que se presentan en las tablas 24 (pag. 117), 25 (pag. 117) y 26 (pag. 118).

Para establecer el criterio de comparación de diseño muestral se calcularon los valores promedio de las diferencias absolutas del estimador con el valor poblacional de los 3 partidos con más votación.

Se observó que los diseños en los que sus estimadores aproximan mejor a los valores poblacionales son: el muestreo aleatorio simple y el muestreo sistemático, lo que se confirma al calcular los promedios para el mejor diseño presentados en la tabla 27 (pag.118).

El diseño de muestreo que mejor funcionó fue el muestreo aleatorio simple de secciones, seguido cercanamente por el muestreo sistemático y supera a los diseños estratificados.

El criterio de estratificación por distrito electoral no consigue influir con claridad en la exactitud para todos los estimadores. Al utilizar distritos electorales o clasificarlos por ubicación urbana/rural de casillas, implícitamente se estratifica geográficamente, lo que no está correlacionado necesariamente con la preferencia electoral de la población.

En particular, al estratificar por distrito electoral y contar con 13, 15 o 18 estratos, la muestra correspondiente para cada estrato es pequeña y los estimadores pierden precisión.

Algunos autores como Cochran han mencionado que las ganancias en precisión son generalmente modestas en la estratificación geográfica; otros como Kish, afirman que las variables más correlacionadas con el voto son: la educación, ocupación y el nivel socioeconómico. Información difícil de recabar en un proceso de elecciones, al no estar permitido conocerlas en el votante, de otra manera se podría relacionar con la del IFE, INEGI y la Secretaría de Hacienda.

Para las encuestas también es difícil obtener con exactitud estos datos, por razones de tipo psicológico, las personas tienden a no declarar con veracidad las preguntas que se les solicitan.

Los resultados de estimación de varianza en diferentes diseños muestrales consiguen un aceptable nivel de similitud con respecto a los valores calculados por fórmula.

Los grupos aleatorios independientes con k=4 obtienen, en general, una mejor aproximación, aunque en algunos casos por la influencia del tamaño de muestra no se observó así.

Fue posible comparar resultados de acuerdo al número de grupos aleatorios y se pudo observar cierta tendencia de disminución del sesgo en algunos casos cuando se incrementa su número, aunque para confirmarlo habría que probarlo con otros tamaño de muestra.

Es posible obtener otras dos estimaciones de varianza al efectuar remuestreo de grupos aleatorios independientes con una muestra y según el diseño. Se calcula la suma de cuadrados de dos maneras; con la media de los grupos aleatorios, o con el estimador del diseño de muestreo.

Los programas diseñados utilizaron rutinas probadas de generación de números aleatorios y manejadores de bases de datos (la información se hallaba en formato compatible) para su almacenamiento. La validación fue realizada al probar las rutinas en una base con datos creados artificialmente, la base contenía el mismo número en todos los registros para el muestreo aleatorio simple, y números iguales en cada estrato para el muestreo estratificado.

Aunque existen otros métodos de estimación de varianza como la linearización de varianza, bootstrap y el jackknife, el método de grupos aleatorios además de su facilidad de implementación consigue resultados aceptables.

ALCANCES Y LIMITACIONES

El tamaño de la muestra escogido para el muestreo aleatorio simple en los estados de Guanajuato, Yucatán y Michoacán del 4%, 9% (con menos secciones) y 6 % respectivamente, se decidió con base en las características de la unidad de muestreo, del nivel de precisión deseado y de los costos de operación. Sería interesante variar los tamaños para estudiar su influencia en la precisión de los estimadores.

Se determinó conservar los tamaños de muestra para los demás diseños con el fin de comparar su comportamiento.

El muestreo estratificado fue menos efectivo al tener muchos estratos, porque la muestra escogida correspondiente es más pequeña en cada uno, pero esto no es la regla general ya que existen diseños de dos unidades por estrato. Además de que el criterio de estratificación no está suficientemente correlacionado con la variable a observar (preferencia de voto), es necesario una muestra más grande y un mejor criterio al estratificar.

Fueron empleados la mayoria de los diseños de muestreo y algunas variantes. Se probaron tres estimadores poco utilizados en el muestreo estratificado de conglomerados, aunque faltó observar el comportamiento del muestreo con probabilidad proporcional al tamaño.

Al utilizar un muestreo con probabilidad proporcional al tamaño de la sección en conteos rápidos, los secciones de la muestra podrian no obtenerse con absoluta seguridad. Las secciones grandes tendrian mayor probabilidad de escogerse, pero menos probabilidad, quizás, de recabarse. Su información se demoraría precisamente por ser de mayor tamaño y no permitirían recabarla a tiempo en un conteo rápido, que por su naturaleza está restringido a un horario determinado. El diseño generaría un mayor retraso de la información o la imposibilidad de cumplir con los resultados oportunamente.

GUANAJUATO

	Estimador	Diseño	Promedio*
1	de razón	Muestreo sistemático	5.25
2	insesgado	Muestreo sistemático	6.00
3	de razón	Muestreo aleatorio simple	6.50
4	insesgado	Muestreo aleatorio simple	9.75
5	insesgado	Muestreo estratificado por distrito	10.50
6	insesgado por estrato	Muestreo estratificado por distrito	56.75
7	de razón por estrato	Muestreo estratificado por distrito	73.50
8	media de medias por estrato	Muestreo estratificado por distrito	172.00
9	media de medias	Muestreo aleatorio simple	173.75
10	media de medias	Muestreo sistemático	179.00

TABLA 24 Valores promedio de diferencias absolutas con respecto al valor poblacional por 10^4 para cada estimador y partido en el estado de Guanajuato. * Promedio sobre PRI, PAN, PRD y Otros.

YUCATÁN

	Estimador	Diseño	Promedio*
1	de razón	Muestreo aleatorio simple	7.25
2	insesgado	Muestreo aleatorio simple	20.50
3	insesgado	Muestreo sistemático	26.50
4	de razón	Muestreo sistemático	27.25
5	insesgado	Muestreo estratificado por distrito	43.50
6	media de medias	Muestreo sistemático	89.75
7	insesgado por estrato	Muestreo estratificado por distrito	93.75
8	de razón por estrato	Muestreo estratificado por distrito	97.75
9	de razón	Muestreo aleatorio simple	100.00
10	media de medias por estrato	Muestreo estratificado por distrito	154.75

TABLA 25 Valores promedio de diferencias absolutas con respecto al valor poblacional por 10^4 para cada estimador y partido en el estado de Yucatán. * Promedio sobre PRI, PAN, PRD y Otros.

MICHOACÁN

	Estimador	Diseño	Promedio*
1	de razón	Muestreo aleatorio simple	10.75
2	insesgado	Muestreo aleatorio simple	11.00
3	insesgado	Muestreo sistemático	16.25
4	de razón	Muestreo sistemático	16.25
5	insesgado	Muestreo estratificado por distrito	16.50
6	insesgado por estrato	Muestreo estratificado por distrito	31.75
7	de razón por estrato	Muestreo estratificado por distrito	34.00
8	media de medias por Estrato	Muestreo estratificado por distrito	118.00
9	media de medias	Muestreo sistemático	120.50
10	media de medias	Muestreo aleatorio simple	135.75

TABLA 26 Valores promedio de diferencias absolutas con respecto al valor poblacional por 10^4 para cada estimador y partido en el estado de Michoacán. * Promedio sobre PRI, PAN, PRD y Otros.

GUANAJUATO		YUCATÁN		MICHOACÁN	
Diseño	Promedio	Diseño	Promedio	Diseño	Promedio
Muestreo aleatorio simple	63.33	Muestreo aleatorio simple	42.48	Muestreo estratificado por distrito	50.06
Muestreo sistemático	63.42	Muestreo sistemático	48.00	Muestreo sistemático	51.11
Muestreo estratificado por distrito	78.19	Muestreo estratificado por distrito	97.44	Muestreo aleatorio simple	52.50

TABLA 27 Valores promedio de diferencias absolutas con respecto al valor poblacional por 10^4 para cada diseño de muestreo en los estados de Guanajuato, Yucatán y Michoacán

Nota aclaratoria:

Debe mencionarse que, para una comparación correcta del método de estimación de la varianza por grupos aleatorios, debió de respetarse formalmente el mismo diseño de muestreo utilizado. Estos es, para obtener los grupos aleatorios dependientes se realizó una partición de la muestra; para los grupos aleatorios independientes se debe reemplazar la muestra obtenida antes de seleccionar la siguiente, sin embargo, el procedimiento aplicado fue escoger uno a uno cada elemento de la muestra con reemplazo. Aún así, se piensa que los resultados que se obtuvieron no deben diferir mucho del método correcto.

BIBLIOGRAFÍA

1. **Azorín Poch, Francisco** . (1972). Curso de Muestreo y Aplicaciones. Aguilar.

2. **Bartley, P., Fox, B.** (1987). A Guide to Simulation. Springer, Nueva York , 2ª edición.

3. **Cochran, W. G.** (1977). Sampling Techniques. John Wiley & Sons, Nueva York, 3ª edición.

4. **Cochran, W. G.** (1946). " Relative accuracy of systematic and stratified random samples for a certain class of populations ".Annals of Mathematical Statistics, Vol. 17, páginas 164-177.

5. **Des Raj**. (1980). Teoria del Muestreo. Fondo de Cultura Económica.

6. **Hansen, M. H., Hurwitz, W. N., Madow, W. G**. (1953). Sample Survey Methods and Theory. John Wiley & Sons, Nueva York, v.1

7. **Kish, L**. (1965). Survey Sampling. John Wiley & Sons, Nueva York.

8. **Kovar, J. G., Rao, J.N.K.,Wu, C. F. J**. (1988) " Bootstrap and other methods to measure errors in survey estimates ". The Canadian Journal of Statistics, Vol. 16, Supplement, páginas 25-45.

9. **Madow , William G**. (1949) "On the Theory of Systematic Sampling II ". Annals of Mathematical Statistics, Vol. 20, páginas 333-354.

10. **Madow , William G.** (1952). " On the Theory of Systematic Sampling III : Comparison of centered and Random Start Systematic Sampling ". University of Illinois.

11. **Madow , William G., Madow Lillian H.** (1944) "On the Theory of Systematic Sampling I " , Annals of Mathematical Statistics, Vol . 15, páginas 1-24.

12. **Madow, Lillian H** .(1946). " Systematic Sampling and its relation to other sampling designs ". American Statistical Association Journal, Vol. 41, páginas 204-207.

13. **Mahalanobis, P. C**. (1946). " Recent Experiments in Statistical Sampling in the Indian Statistical Institute " . Journal of the Royal Statistical Society, Vol. CIX., Part. IV. Londres, Inglaterra.

14. **Stephan, F.F., McCarthy, P. J.** (1958). Sampling Opinions: an analysis of Survey Procedure. John Wiley & Sons, Nueva York.

15. **Sukhatme, Pandurang V., Sukhatme, Balkrishna V., Sukhatme, Shashikala., Asok, C.** (1984). Sampling Theory of Surveys, with Applications. Iowa State College Press, Ames Iowa, 3ª edición.

16. **Wolter, Kirk M.** (1985). Introduction to Variance Estimation. Springer-Verlag, Nueva York.

ANEXO . Programas de Cómputo

Se presentan los programas fuente utilizados para la simulación de conteos rápidos en el Estado de Michoacán, para los demás Estados se utiliza un procedimiento análogo. Se describen las equivalencias computacionales para cada estimador

Se usa el lenguaje de programación CLIPPER versión 5.2 que al compilar utiliza rutinas de soporte llamadas 'Microsoft C floating Point Support Routines'.

A continuación se detalla un índice de los programas

P 1 (PROGRAMA 1 : MUESTRA ALEATORIA SIMPLE SIN REEMPLAZO)
ENTRADA : P=TAMAÑO DE POBLACIÓN, Q=TAMAÑO DE MUESTRA,
B=NOMBRE DE LA BASE DE DATOS

SALIDA : BASE DE DATOS VALIDO.DBF CON Q ELEMENTOS MUESTRA SIN REEMPLAZO

```
************************************************
*PROCEDIMIENTO QUE UTILIZA UN ALGORITMO PARA OBTENER MUESTRA SIN REEMPLAZO
* GENERADA POR UNA SEMILLA .
*TAMAÑO DE POBLACIÓN " P " , TAMAÑO DE MUESTRA " Q ", "B" NOMBRE DE LA BASE DE
*DATOS
*TOMADO DE " BARTLEY, P. ,A GUIDE TO SIMULATION, 1987,2 ED."
* EL CUAL GENERA NÚMEROS ALEATORIOS CON REEMPLAZO
************************************************
PROCEDURE MUES(P,Q,B)                 && Procedimiento MUES(MUEstras Sin reemplazo)
 B=UPPER(TRIM(B))+".DBF "             && calcula una muestra sin reemplazo
 SET DECIMALS TO 10
  USE VALIDO                          && utiliza la base de validación VALIDO
  ZAP                                 && borra los registros (parte de cero)
  USE
  CLEAR
  RESTORE FROM SEMILLA ADDITIVE       && Agrega a las variables de entorno la semilla
                                      && previamente grabada
  IF SEMILLA=2147483647               && si la semilla alcanza su máximo,
     SEMILLA=0                        && se incializa en 1
  ENDIF
  SEMILLA=SEMILLA+1                   && de otra forma se incrementa en 1
  SEMILLA=INT(SEMILLA)
  IM=SEMILLA
  IML=IM                              && parte del algoritmo
  I=1
**********
  ? "BASE DE DATOS : "+B                     && se presenta en pantalla los datos de el
  ? "SEMILLA : "+LTRIM(STR(SEMILLA))         && nombre de la base de datos, la semilla
  ?"TAMAÑODE POBLACION : "+LTRIM(STR(P))  && el tamaño de la población, de muestra
  ? "TAMAÑO DE MUESTRA   : "+LTRIM(STR(Q))
  ?
 DO WHILE I<=Q                  && generar hasta completar la muestra números aleatorios
   K1=IM/127773
   K1=LTRIM(STR(K1))
   K1=SUBSTR(K1,1,6)
   K1=VAL(K1)
   IM=(16807*(IM-(K1*127773)))-(K1*2836)
   IM2=IM

    IF IM <0
      IM=IM+2147483647
   ENDIF
   UNIF=IM*0.0000000004656612875
   U=UNIF*P*10000
  SET DECIMALS TO 0
```

```
 U=ROUND(U,0)                        && el algoritmo genera un número aleatorio U

IF U <=P.AND.U>0                      && si todavía no se alcanza el tamaño de muestra
 USE VALIDO INDEX NUMERO
  REINDEX
SEEK U                                && busca el número aleatorio U
  IF .NOT. FOUND()                    && si no lo encuentra
    APPEND RECORD U FROM &B           &&  lo agrega a la muestra
    REPLACE NUMERO WITH U              && reemplaza al campo NUMERO con U
   I=I+1
  ENDIF

ENDIF
SET DECIMALS TO 10
ENDDO
SAVE TO SEMILLA ALL LIKE SEMILLA*  && Al final salva la semilla para que se utilice una nueva
SET DECIMALS TO 4                     && la próxima vez
RETURN
```

P 2 (PROGRAMA 2 : MUESTRA ALEATORIA SIMPLE CON REEMPLAZO)

ENTRADA : P=TAMAÑO DE POBLACIÓN, Q=TAMAÑO DE MUESTRA,
B=NOMBRE DE LA BASE DE DATOS

SALIDA : BASE DE DATOS VALIDO.DBF CON Q ELEMENTOS MUESTRA CON REEMPLAZO

```
***************************************************
*ALGOC.PRG
*PROCEDIMIENTO QUE UTILIZA UN ALGORITMO PARA OBTENER MUESTRA CON
*REEMPLAZO GENERADA POR UNA SEMILLA .
*TAMAÑO DE POBLACION " P " Y TAMAÑO DE MUESTRA " Q "
*TOMADO DE " BARTLEY. P. ,A GUIDE TO SIMULATION, 1987,2 ED "
* EL CUAL GENERA NUMEROS ALEATORIOS CON REPETICIÓN
***************************************************
PROCEDURE MUEC(P,Q,B)  && Procedimiento MUEC(MUEstra Con reemplazo)
B=UPPER(TRIM(B))+".DBF"   && básicamente es análogo a el procedimiento de
SET DECIMALS TO 10        && Muestras sin Reemplazo
USE VALIDO
ZAP
USE
CLEAR
RESTORE FROM SEMILLA ADDITIVE
IF SEMILLA=2147483647
 SEMILLA=0
ENDIF
SEMILLA=SEMILLA+1
SEMILLA=INT(SEMILLA)
IM=SEMILLA
IML=IM
I=1
Q=ROUND(Q.0)
```

```
Q=INT(Q)

? "BASE DE DATOS : "+B
? "SEMILLA : "+LTRIM(STR(SEMILLA))
? "TAMAÑO DE POBLACION : "+LTRIM(STR(P))
? "TAMAÑO DE MUESTRA   : "+LTRIM(STR(Q))
?
DO WHILE I<=Q
 K1=IM/127773
 K1=LTRIM(STR(K1))
 K1=SUBSTR(K1,1,6)
 K1=VAL(K1)
 IM=(16807*(IM-(K1*127773)))-(K1*2836)
 IM2=IM

 IF IM <0
   IM=IM+2147483647
 ENDIF
 UNIF=IM*0.0000000004656612875
 P=INT(P)
U=UNIF*P*10000
 SET DECIMALS TO 0
U=INT(ROUND(U,0))
IF U <=P.AND.U>0
  USE VALIDO                    && utiliza la base VALIDO y permite la repitición de registros
APPEND RECORD U FROM &B

 I =I+1
ENDIF
SET DECIMALS TO 10
ENDDO
SAVE TO SEMILLA ALL LIKE SEMILLA*
SET DECIMALS TO 4
RETURN
```

P 3 (PROGRAMA 3 : ESTIMADORES CON UNA MUESTRA ALEATORIA SIMPLE SIN REEMPLAZO DE TAMAÑO 125 (6%) DE LA BASE MICH.DBF DEL ESTADO DE MICHOACAN CON 2081 SECCIONES , GENERA GRUPOS ALEATORIOS INDEPENDIENTES Y DEPENDIENTES CON K=2,4)

```
***************************
CASO="SIN" && EL CASO ES MUESTRA ALEATORIA SIMPLE POR OMISIÓN, PERO PUEDE
*MODIFICARSE PARA QUE LA VARIABLE SEA PARA UN MUESTREO SISTEMÁTICO
* MEDIANTE: CASO="SIS"

H=1
QV=1
CLEAR
T=SECONDS()
?
? LTRIM(STR(H))+" DE "+ LTRIM(STR(QV))
USE
```

```
MB=426        && PROMEDIO DE LAS Mi (VALIDOS)
CLS
SET COLOR TO W/B,B/W,B
SET SAFETY OFF
SET DECIMALS TO 0
SET EXACT ON
USE VALIDO
TER="NO"
ZAP
IF CASO="SIN"
MUES(2081,125,"MICH") && MUESTRA SIN REEMPLAZO DE MICH.DBF
ENDIF
ENDIF
**************************************************
IF TER="NO"
CLEAR


USE VALIDO
USE
!COPY VALIDO.DBF MAS1.DBF      && Copia la muestra auna base llamada MAS1.DBF
SET DECIMAL TO 4              && para utilizarla posteriormente en la generación de Grupos A.
USE VALIDO    && Utilizamos la muestra
```

SUM PRI TO SPRI && $SPRI = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} M_i \bar{y}_i.$ suma de los votos para el PRI de las secciones

&& elegidas en la muestra

```
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
```

SUM VALIDOS TO TTOTAL && $TTOTAL = \sum_{i=1}^{n} M_i$ suma de los tamaños de las secciones de la

&& muestra

```
AVERAGE PRI TO E2PRI
```

E2PRI=E2PRI/MB && $E2PRI = \frac{\sum_{i=1}^{n} y_i}{n\overline{M}} = \frac{\sum_{i=1}^{n} M_i \bar{y}_i.}{n\overline{M}}$ El estimador insesgado

```
AVERAGE PAN TO E2PAN
E2PAN=E2PAN/MB
AVERAGE PRD TO E2PRD
E2PRD=E2PRD/MB
AVERAGE OTROS TO E2OTROS
E2OTROS=E2OTROS/MB
```

AVERAGE PRI/VALIDOS TO E1PRI && $E1PRI = \frac{\sum_{i=1}^{n} \bar{y}_i.}{n}$ el estimador media de medias

```
AVERAGE PAN/VALIDOS TO E1PAN
AVERAGE PRD/VALIDOS TO E1PRD
AVERAGE OTROS/VALIDOS TO E1OTROS
```

```
AVERAGE VALIDOS TO PTOTAL
*********** EL ESTIMADOR DE RAZON :
```

PPRI=SPRI/TTOTAL && PPRI= $\dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \dfrac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i}$ el estimador de razón

```
PPAN=SPAN/TTOTAL
PPRD=SPRD/TTOTAL
POTROS=SOTROS/TTOTAL
IF CASO="SIS"
M=125
ENDIF
ENDIF
USE RESULME2
APPEND BLANK
IF CASO="SIS"
REPLACE ARRANQ WITH NUMER
ENDIF
REPLACE RPRI1 WITH E1PRI,RPAN1 WITH E1PAN  && enviamos una base de datos los
                                           && resultados
REPLACE RPRD1 WITH E1PRD,ROTROS1 WITH E1OTROS
REPLACE RPRI2 WITH E2PRI,RPAN2 WITH E2PAN
REPLACE RPRD2 WITH E2PRD,ROTROS2 WITH E2OTROS
REPLACE RPRI3 WITH PPRI,RPAN3 WITH PPAN
REPLACE RPRD3 WITH PPRD,ROTROS3 WITH POTROS

DO GA2    && Calcular grupos aleatorios  independientes k=2
DO GA4    && Calcular grupos aleatorios  independientes k=4
SET DECIMALS TO 0
DO GAD2   && Calcular grupos aleatorios  dependientes k=2
DO GAD4   && Calcular grupos aleatorios  dependientes k=4
? TSTRING(SECONDS()-T)
#INCLUDE "ALGOS.PRG"  && utiliza los procedimientos de muestra sin y con
#INCLUDE "ALGOC.PRG"  && reemplazo
```

P 4 (PROGRAMA 4 : GRUPOS ALEATORIOS INDEPENDIENTES CON K=2 DE UNA MUESTRA ALEATORIA SIMPLE SIN REEMPLAZO PARA DOS ESTIMADORES, ASÍ COMO SE CALCULAN LOS ESTIMADORES PARA LAS VARIANZAS)

```
************************
* GA2.PRG
USE VALIDO
ZAP
CASO="CON"
CLS
SET COLOR TO W/B,B/W,B
SET SAFETY OFF
SET DECIMALS TO 0
SET EXACT ON
```

```
CASOK=2
SET DECIMALS TO 0
SET EXACT ON
CASOA=1
DO WHILE CASOA<=CASOK   && se generan 2 grupos aleatorios independientes
 CASOR=LTRIM(STR(CASOA))
 CASOG=1
 DO WHILE CASOG<=1
  CASOC=LTRIM(STR(CASOG))
  USE MAS&CASOC
  CASON=RECCOUNT()
  USE
  @ 2,3 SAY " GRUPOS A. INDEP.  K=2   "
  @ 2, 32 say CASON
  @ 3,32 SAY CASOG
  N=CASON
  M=CASOK
  M=ROUND((1/M)*N,0)
MUEC(N,M,"MAS&CASOC")   && Muestra con  reemplazo de tamaño M/2 de la base MAS1.DBF
  CASOG=CASOG+1         && previamente generado por un Muestreo Aleatorio Simple
ENDDO
CASOA=CASOA+1

CASOR=CASOR+".DBF"
USE
!COPY VALIDO.DBF BA&CASOR  && se copian el resultado de las muestras a 2 bases de datos
                           && BA1.DBF y BA2.DBF
USE VALIDO
ZAP
ENDDO
*********
SET DECIMALS TO 8
USE GA2
ZAP
C=1
CC=LTRIM(STR(C))

DO WHILE C<=CASOK      && De las bases generadas se evalúan los estimadores que sus valores
                       && dependen únicamente de la muestra , el media de medias y el de razón
USE BA&CC
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM VALIDOS TO SVAL

AVERAGE (PRI/VALIDOS) TO MBA&CC    && MBA&CC = (sum_{i=1}^{n} ybar_i) / n  estimador media de medias

AVERAGE (PAN/VALIDOS) TO MBB&CC
AVERAGE (PRD/VALIDOS) TO MBC&CC
```

$$MBA\&CC = \frac{\sum_{i=1}^{n} \bar{y}_i}{n}$$

```
MBA2&CC=SPRI/SVAL
```

&& $MBA2\&CC = \dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \dfrac{\sum_{i=1}^{n} M_i \bar{y}_i.}{\sum_{i=1}^{n} M_i}$ estimador de razón

```
MBB2&CC=SPAN/SVAL
MBC2&CC=SPRD/SVAL
USE GA2                     && Se guardan los resultados de los estimadores en la base GA2.DBF
APPEND BLANK
REPLACE MYTI WITH MBA&CC
REPLACE MYTI2 WITH MBA2&CC
REPLACE MYTI3 WITH MBB&CC
REPLACE MYTI4 WITH MBB2&CC
REPLACE MYTI5 WITH MBC&CC
REPLACE MYTI6 WITH MBC2&CC
C=C+1
CC=LTRIM(STR(C))
ENDDO
***
USE GA2
AVERAGE MYTI TO ME       && Se promedian los resultados de los estimadores de la media
AVERAGE MYTI2 TO ME2
AVERAGE MYTI3 TO ME3
AVERAGE MYTI4 TO ME4
AVERAGE MYTI5 TO ME5
AVERAGE MYTI6 TO ME6
GO TOP
SUM (MYTI-ME)**2 TO SC
SUM (MYTI2-ME2)**2 TO SC2
SUM (MYTI3-ME3)**2 TO SC3
SUM (MYTI4-ME4)**2 TO SC4
SUM (MYTI5-ME5)**2 TO SC5
SUM (MYTI6-ME6)**2 TO SC6
VAG= SC/((CASOK)*(CASOK-1))
VAG2= SC2/((CASOK)*(CASOK-1))
VAG5= SC3/((CASOK)*(CASOK-1))
VAG6= SC4/((CASOK)*(CASOK-1))
VAG9= SC5/((CASOK)*(CASOK-1))
VAG10= SC6/((CASOK)*(CASOK-1))    && el estimador de varianza de Grupos Aleatorios para cada
                                   && estimador es obtenido :
```

&& $VAG = v(\hat{\bar{\theta}}) = \dfrac{\sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\bar{\theta}})^2}{k(k-1)}$

```
MEG=ME
MEG2=ME2
MEG5=ME3
MEG6=ME4
MEG9=ME5
MEG10=ME6
```

```
USE VAGS            && finalmente se envian a una base de datos de resultados llamada
APPEND BLANK     && VAGS.DBF donde cada estimador ocupa un lugar determinado
REPLACE CMEG WITH MEG
REPLACE CVAG WITH VAG
REPLACE CMEG2 WITH MEG2
REPLACE CVAG2 WITH VAG2

REPLACE CMEG5 WITH MEG5
REPLACE CVAG5 WITH VAG5
REPLACE CMEG6 WITH MEG6
REPLACE CVAG6 WITH VAG6

REPLACE CMEG9 WITH MEG9
REPLACE CVAG9 WITH VAG9
REPLACE CMEG10 WITH MEG10
REPLACE CVAG10 WITH VAG10

SET DECIMALS TO 4
```

P 5 (PROGRAMA 5 : GRUPOS ALEATORIOS INDEPENDIENTES CON K=4 DE UNA MUESTRA ALEATORIA SIMPLE SIN REEMPLAZO, ASÍ COMO SE CALCULAN LOS ESTIMADORES PARA LAS VARIANZAS)

```
*****************************
* GA4.PRG
USE VALIDO
ZAP
CASO="CON"
CLS
SET COLOR TO W/B,B/W,B
SET SAFETY OFF
SET DECIMALS TO 0
SET EXACT ON
CASOK=4
SET DECIMALS TO 0
SET EXACT ON
CASOA=1
DO WHILE CASOA<=CASOK
CASOR=LTRIM(STR(CASOA))

CASOG=1
DO WHILE CASOG<=1
CASOC=LTRIM(STR(CASOG))
USE MAS&CASOC
CASON=RECCOUNT()
USE

@ 2,3 SAY " GRUPOS A. INDEP.  K=4   "
@ 2, 32 say CASON
@ 3,32 SAY CASOG
N=CASON
M=CASOK
```

```
M=ROUND((1/M)*N,0)
MUEC(N,M,"MAS&CASOC")
CASOG=CASOG+1
ENDDO
CASOA=CASOA+1
CASOR=CASOR+".DBF"
USE
!COPY VALIDO.DBF BA&CASOR
USE VALIDO
ZAP
ENDDO

SET DECIMALS TO 8
USE GA2
ZAP
C=1
CC=LTRIM(STR(C))

DO WHILE C<=CASOK
USE BA&CC
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM VALIDOS TO SVAL
AVERAGE (PRI/VALIDOS) TO MBA&CC
AVERAGE (PAN/VALIDOS) TO MBB&CC
AVERAGE (PRD/VALIDOS) TO MBC&CC
MBA2&CC=SPRI/SVAL
MBB2&CC=SPAN/SVAL
MBC2&CC=SPRD/SVAL
USE GA2
APPEND BLANK
REPLACE MYTI WITH MBA&CC
REPLACE MYTI2 WITH MBA2&CC
REPLACE MYTI3 WITH MBB&CC
REPLACE MYTI4 WITH MBB2&CC
REPLACE MYTI5 WITH MBC&CC
REPLACE MYTI6 WITH MBC2&CC
C=C+1
CC=LTRIM(STR(C))
ENDDO
***
USE GA2
AVERAGE MYTI TO ME
AVERAGE MYTI2 TO ME2
AVERAGE MYTI3 TO ME3
AVERAGE MYTI4 TO ME4
AVERAGE MYTI5 TO ME5
AVERAGE MYTI6 TO ME6
GO TOP
SUM (MYTI-ME)**2 TO SC
SUM (MYTI2-ME2)**2 TO SC2
```

```
SUM (MYTI3-ME3)**2 TO SC3
SUM (MYTI4-ME4)**2 TO SC4
SUM (MYTI5-ME5)**2 TO SC5
SUM (MYTI6-ME6)**2 TO SC6

VAG= SC/((CASOK)*(CASOK-1))
VAG2= SC2/((CASOK)*(CASOK-1))
VAG5= SC3/((CASOK)*(CASOK-1))
VAG6= SC4/((CASOK)*(CASOK-1))
VAG9= SC5/((CASOK)*(CASOK-1))
VAG10= SC6/((CASOK)*(CASOK-1))


MEG=ME
MEG2=ME2
MEG5=ME3
MEG6=ME4
MEG9=ME5
MEG10=ME6

USE VAG2S
APPEND BLANK
REPLACE CMEG WITH MEG
REPLACE CVAG WITH VAG
REPLACE CMEG2 WITH MEG2
REPLACE CVAG2 WITH VAG2

REPLACE CMEG5 WITH MEG5
REPLACE CVAG5 WITH VAG5
REPLACE CMEG6 WITH MEG6
REPLACE CVAG6 WITH VAG6

REPLACE CMEG9 WITH MEG9
REPLACE CVAG9 WITH VAG9
REPLACE CMEG10 WITH MEG10
REPLACE CVAG10 WITH VAG10

SET DECIMALS TO 4
```

P 6 (PROGRAMA 6 : GRUPOS ALEATORIOS DEPENDIENTES CON K=2 DE UNA MUESTRA ALEATORIA SIMPLE SIN REEMPLAZO PARA DOS ESTIMADORES, ASÍ COMO SE CALCULAN LOS ESTIMADORES PARA LAS VARIANZAS)

```
***************
*GAD2.PRG                                   && Una de la maneras de obtener Grupos Aleatorios
GRUPOS ALEATORIOS DEPENDIENTES (PARES E IMPARES)  && Dependientes es la siguiente:
USE PAR
ZAP                                         && Particionando la base de datos en dos grupos:
USE IMP                                     && Pares e Impares
ZAP

SET FIXED ON
CLEAR
```

```
SET DECIMALS TO 8
SET TALK OFF
E=1
J=1
I=1
C=LTRIM(STR(J))

DO WHILE J<=1


 USE MAS&C
 UL=RECCOUNT()
 GO TOP
 DO WHILE .NOT. EOF()
  REG=RECNO()
     CA=PRI/VALIDOS
     CA2=PRI
     CA3=VALIDOS
     CA4=PAN/VALIDOS
     CA5=PAN
     CA6=CA3
     CA7=PRD/VALIDOS
     CA8=PRD
     CA9=CA3
     USE IMP
     APPEND BLANK
     REPLACE CIMP WITH CA,CIMP2 WITH CA2
     REPLACE CIMP3 WITH CA3,CIMP4 WITH CA4
     REPLACE CIMP5 WITH CA5,CIMP6 WITH CA6
     REPLACE CIMP7 WITH CA7,CIMP8 WITH CA8
     REPLACE CIMP9 WITH CA9
     USE MAS&C

  IF REG+1>UL
   EXIT
  ENDIF
GO REG+1
     CA=PRI/VALIDOS
     CA2=PRI
     CA3=VALIDOS
     CA4=PAN/VALIDOS
     CA5=PAN
     CA6=CA3
     CA7=PRD/VALIDOS
     CA8=PRD
     CA9=CA3

    USE PAR
     APPEND BLANK
     REPLACE CPAR WITH CA,CPAR2 WITH CA2
     REPLACE CPAR3 WITH CA3
     REPLACE CPAR4 WITH CA4
```

```
    REPLACE CPAR5 WITH CA5,CPAR6 WITH CA6
    REPLACE CPAR7 WITH CA7,CPAR8 WITH CA8
    REPLACE CPAR9 WITH CA9

   USE MAS&C
  IF REG+2>UL
    EXIT
  ENDIF
 GO REG+2
ENDDO

J=J+1
C=LTRIM(STR(J))
ENDDO

*HS2D.PRG
SET DECIMALS TO 8

USE PAR
SUM CPAR2 TO SPRI1
SUM CPAR3 TO SVAL1
```

MP2=SPRI1/SVAL1 && $MP2 = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n} M_i \bar{y}_{i\cdot}}{\sum_{i=1}^{n} M_i}$ estimador de razón

AVERAGE CPAR TO MP && $MP = \frac{\sum_{i=1}^{n} \bar{y}_{i\cdot}}{n}$ estimador media de medias

```
SUM CPAR5 TO SPAN1
SUM CPAR6 TO SVAL1B
MP2B=SPAN1/SVAL1B
AVERAGE CPAR4 TO MPB

SUM CPAR8 TO SPRD1
SUM CPAR9 TO SVAL1C
MP2C=SPRD1/SVAL1C
AVERAGE CPAR7 TO MPC


USE IMP
SUM CIMP2 TO SPRI2
SUM CIMP3 TO SVAL2
MI2=SPRI2/SVAL2
AVERAGE CIMP TO MI

SUM CIMP5 TO SPAN2
SUM CIMP6 TO SVAL2B
```

```
MI2B=SPAN2/SVAL2B
AVERAGE CIMP4 TO MIB

SUM CIMP8 TO SPRD2
SUM CIMP9 TO SVAL2C
MI2C=SPRD2/SVAL2C
AVERAGE CIMP7 TO MIC

M1=(MP+MI)/2
M1B=(MPB+MIB)/2
M1C=(MPC+MIC)/2
M2=(MP2+MI2)/2
M2B=(MP2B+MI2B)/2
M2C=(MP2C+MI2C)/2
EVM=(((MP-M1)**2)+((MI-M1)**2))/2
EVMB=(((MPB-M1B)**2)+((MIB-M1B)**2))/2
EVMC=(((MPC-M1C)**2)+((MIC-M1C)**2))/2
EVM2=(((MP2-M2)**2)+((MI2-M2)**2))/2
EVM2B=(((MP2B-M2B)**2)+((MI2B-M2B)**2))/2
EVM2C=(((MP2C-M2C)**2)+((MI2C-M2C)**2))/2
USE VAGS

GO BOTTOM
REPLACE CMEG3 WITH M1
REPLACE CVAG3 WITH EVM
REPLACE CMEG4 WITH M2
REPLACE CVAG4 WITH EVM2

REPLACE CMEG7 WITH M1B
REPLACE CVAG7 WITH EVMB
REPLACE CMEG8 WITH M2B
REPLACE CVAG8 WITH EVM2B

REPLACE CMEG11 WITH M1C
REPLACE CVAG11 WITH EVMC
REPLACE CMEG12 WITH M2C
REPLACE CVAG12 WITH EVM2C

SET DECIMALS TO 4
```

P 7 (PROGRAMA 7 : GRUPOS ALEATORIOS DEPENDIENTES CON K=4 DE UNA MUESTRA ALEATORIA SIMPLE SIN REEMPLAZO PARA DOS ESTIMADORES, ASÍ COMO SE CALCULAN LOS ESTIMADORES PARA LAS VARIANZAS)
****

GAD4.PRG

```
! COPY GANI.DBF GNI1.DBF
! COPY GANI.DBF GNI2.DBF
! COPY GANI.DBF GNI3.DBF
! COPY GANI.DBF GNI4.DBF
```

```
SET FIXED ON
CLEAR
SET DECIMALS TO 8
SET TALK OFF
E=1
J=1
I=1
C=LTRIM(STR(J))

DO WHILE J<=1

 USE MAS&C
 UL=RECCOUNT()
 GO TOP
DO WHILE .NOT. EOF()
 REG=RECNO()
    CA=PRI/VALIDOS
    CA2=PRI
    CA3=VALIDOS
    CA4=PAN/VALIDOS
    CA5=PAN
    CA6=CA3
    CA7=PRD/VALIDOS
    CA8=PRD
    CA9=CA3



   USE GN11
    APPEND BLANK
    REPLACE CAM WITH CA,CAM2 WITH CA2
    REPLACE CAM3 WITH CA3,CAM4 WITH CA4
    REPLACE CAM5 WITH CA5,CAM6 WITH CA6
    REPLACE CAM7 WITH CA7, CAM8 WITH CA8
    REPLACE CAM9 WITH CA9
  USE MAS&C

 IF REG+1>UL
  EXIT
 ENDIF
GO REG+1
CA=PRI/VALIDOS
    CA2=PRI
    CA3=VALIDOS
    CA4=PAN/VALIDOS
    CA5=PAN
    CA6=CA3
    CA7=PRD/VALIDOS
    CA8=PRD
    CA9=CA3
```

```
USE GNI2
    APPEND BLANK
    REPLACE CAM WITH CA.CAM2 WITH CA2
    REPLACE CAM3 WITH CA3,CAM4 WITH CA4
    REPLACE CAM5 WITH CA5,CAM6 WITH CA6
    REPLACE CAM7 WITH CA7. CAM8 WITH CA8
    REPLACE CAM9 WITH CA9

  USE MAS&C
  IF REG+2>UL
   EXIT
  ENDIF
 GO REG+2
CA=PRI/VALIDOS
    CA2=PRI
    CA3=VALIDOS
    CA4=PAN/VALIDOS
    CA5=PAN
    CA6=CA3
    CA7=PRD/VALIDOS
    CA8=PRD
    CA9=CA3

   USE GNI3
    APPEND BLANK
    REPLACE CAM WITH CA,CAM2 WITH CA2
    REPLACE CAM3 WITH CA3.CAM4 WITH CA4
    REPLACE CAM5 WITH CA5,CAM6 WITH CA6
    REPLACE CAM7 WITH CA7. CAM8 WITH CA8
    REPLACE CAM9 WITH CA9


  USE MAS&C
  IF REG+3>UL
   EXIT
  ENDIF
  GO REG+3
  CA=PRI/VALIDOS
    CA2=PRI
    CA3=VALIDOS
    CA4=PAN/VALIDOS
    CA5=PAN
    CA6=CA3
    CA7=PRD/VALIDOS
    CA8=PRD
    CA9=CA3

    USE GNI4
    APPEND BLANK
    REPLACE CAM WITH CA.CAM2 WITH CA2
    REPLACE CAM3 WITH CA3,CAM4 WITH CA4
    REPLACE CAM5 WITH CA5,CAM6 WITH CA6
```

```
    REPLACE CAM7 WITH CA7, CAM8 WITH CA8
    REPLACE CAM9 WITH CA9


   USE MAS&C
  IF REG+4>UL
   EXIT
  ENDIF
 GO REG+4

 ENDDO

J=J+1
C=LTRIM(STR(J))
ENDDO


*HS42.PRG
SET DECIMALS TO 8
USE GNI1

SUM CAM2 TO SPRI1
SUM CAM3 TO SVAL1
MEDR1=SPRI1/SVAL1
AVERAGE CAM TO MED1

SUM CAM5 TO SPAN1
SUM CAM6 TO SVAL1B
MEDR1B=SPAN1/SVAL1B
AVERAGE CAM4 TO MED1B
SUM CAM8 TO SPRD1
SUM CAM9 TO SVAL1C
MEDR1C=SPRD1/SVAL1C
AVERAGE CAM7 TO MED1C

USE GNI2
SUM CAM2 TO SPRI2
SUM CAM3 TO SVAL2
MEDR2=SPRI2/SVAL2
AVERAGE CAM TO MED2

SUM CAM5 TO SPAN2
SUM CAM6 TO SVAL2B
MEDR2B=SPAN2/SVAL2B
AVERAGE CAM4 TO MED2B

SUM CAM8 TO SPRD2
SUM CAM9 TO SVAL2C
MEDR2C=SPRD2/SVAL2C
AVERAGE CAM7 TO MED2C

USE GNI3
```

```
SUM CAM2 TO SPRI3
SUM CAM3 TO SVAL3
MEDR3=SPRI3/SVAL3
AVERAGE CAM TO MED3

SUM CAM5 TO SPAN3
SUM CAM6 TO SVAL3B
MEDR3B=SPAN3/SVAL3B
AVERAGE CAM4 TO MED3B

SUM CAM8 TO SPRD3
SUM CAM9 TO SVAL3C
MEDR3C=SPRD3/SVAL3C
AVERAGE CAM7 TO MED3C

USE GNI4
SUM CAM2 TO SPRI4
SUM CAM3 TO SVAL4
MEDR4=SPRI4/SVAL4
AVERAGE CAM TO MED4

SUM CAM5 TO SPAN4
SUM CAM6 TO SVAL4B
MEDR4B=SPAN4/SVAL4B
AVERAGE CAM4 TO MED4B

SUM CAM8 TO SPRD4
SUM CAM9 TO SVAL4C
MEDR4C=SPRD4/SVAL4C
AVERAGE CAM7 TO MED4C
M1=(MED1+MED2+MED3+MED4)/4
M1B=(MED1B+MED2B+MED3B+MED4B)/4
M1C=(MED1C+MED2C+MED3C+MED4C)/4
EVM=(((MED1-M1)**2)+((MED2-M1)**2)+((MED3-M1)**2)+((MED4-M1)**2))/(4*3)
EVMB=(((MED1B-M1B)**2)+((MED2B-M1B)**2)+((MED3B-M1B)**2)+((MED4B-M1B)**2))/(4*3)
EVMC=(((MED1C-M1C)**2)+((MED2C-M1C)**2)+((MED3C-M1C)**2)+((MED4C-M1C)**2))/(4*3)

M2=(MEDR1+MEDR2+MEDR3+MEDR4)/4
M2B=(MEDR1B+MEDR2B+MEDR3B+MEDR4B)/4
M2C=(MEDR1C+MEDR2C+MEDR3C+MEDR4C)/4

EVM2=(((MEDR1-M2)**2)+((MEDR2-M2)**2)+((MEDR3-M2)**2)+((MEDR4-M2)**2))/(4*3)
EVM2B=(((MEDR1B-M2B)**2)+((MEDR2B-M2B)**2)+((MEDR3B-M2B)**2)+((MEDR4B-
M2B)**2))/(4*3)
EVM2C=(((MEDR1C-M2C)**2)+((MEDR2C-M2C)**2)+((MEDR3C-M2C)**2)+((MEDR4C-
M2C)**2))/(4*3)

USE VAG2S
 GO BOTTOM
 REPLACE CMEG3 WITH M1
 REPLACE CVAG3 WITH EVM
 REPLACE CMEG4 WITH M2
```

```
REPLACE CVAG4 WITH EVM2

REPLACE CMEG7 WITH M1B
REPLACE CVAG7 WITH EVMB
REPLACE CMEG8 WITH M2B
REPLACE CVAG8 WITH EVM2B

REPLACE CMEG11 WITH M1C
REPLACE CVAG11 WITH EVMC
REPLACE CMEG12 WITH M2C
REPLACE CVAG12 WITH EVM2C

SET DECIMALS TO 4
```

P 8 (PROGRAMA 8 : ESTIMADORES CON UNA MUESTRA SISTEMÁTICA DE TAMAÑO 125 (6%) DE LA BASE MICH.DBF DEL ESTADO DE MICHOACÁN CON 2081 SECCIONES , K= 17)

```
*P8.PRG
H=1
QV=1
CLEAR
T=SECONDS()

USE
MB=426         && PROMEDIO DE LAS Mi (VALIDOS)
CASO="SIS"
CLS
SET COLOR TO W/B,B/W,B
SET SAFETY OFF
SET DECIMALS TO 0
SET EXACT ON
USE VALIDO
TER="NO"
ZAP
MB=426
IF CASO="SIS"
MUES(17,1,"MICH")
N=17   &&
 M=1
ENDIF
IF TER="NO"
CLEAR
IF CASO="SIS"
USE VALIDO
? "         ARRANQUE ALEATORIO = "+LTRIM(STR(NUMERO))
? "      GENERANDO MUESTRA SISTEMATICA"
STORE NUMERO TO NUM
NUMER=NUM
P=17
W=1
USE VALIDO2
ZAP
DO WHILE W<=125 && TENIA 183
```

```
APPEND RECORD NUM FROM MICH.DBF
NUM=NUM+P
 IF NUM>=2084
   NUM=1
 ENDIF
W=W+1
ENDDO
*R=R+1
*      SAVE TO REGIS.MEM ALL EXCEPT CASO*
*USE VALIDO2
ELSE
USE VALIDO
ENDIF
USE
!COPY VALIDO2.DBF MAS1.DBF
SET DECIMAL TO 4
USE VALIDO2
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
SUM VALIDOS TO TTOTAL
AVERAGE PRI TO E2PRI && ESTIMADOR 2
E2PRI=E2PRI/MB
AVERAGE PAN TO E2PAN
E2PAN=E2PAN/MB
AVERAGE PRD TO E2PRD
E2PRD=E2PRD/MB
AVERAGE OTROS TO E2OTROS
E2OTROS=E2OTROS/MB

AVERAGE PRI/VALIDOS TO E1PRI && ESTIMADOR 1
AVERAGE PAN/VALIDOS TO E1PAN
AVERAGE PRD/VALIDOS TO E1PRD
AVERAGE OTROS/VALIDOS TO E1OTROS
AVERAGE VALIDOS TO PTOTAL
PPRI=SPRI/TTOTAL && ESTIMADOR 3
PPAN=SPAN/TTOTAL
PPRD=SPRD/TTOTAL
POTROS=SOTROS/TTOTAL

ENDIF
USE RESULMS
APPEND BLANK
IF CASO="SIS"
REPLACE ARRANQ WITH NUMER
ENDIF
REPLACE RPRI1 WITH E1PRI,RPAN1 WITH E1PAN
REPLACE RPRD1 WITH E1PRD,ROTROS1 WITH E1OTROS
REPLACE RPRI2 WITH E2PRI,RPAN2 WITH E2PAN
REPLACE RPRD2 WITH E2PRD,ROTROS2 WITH E2OTROS
REPLACE RPRI3 WITH PPRI,RPAN3 WITH PPAN
```

```
REPLACE RPRD3 WITH PPRD,ROTROS3 WITH POTROS
DO GA2S  && OPERA DE LA MISMA FORMA QUE GA2 PERO EL RESULTADO ES ENVIADO
         && A OTRA BASE DE DATOS
DO GA4S
SET DECIMALS TO 0
DO GAD2S  && OPERA DE LA MISMA FORMA QUE GA2 PERO EL RESULTADO ES
          &&  ENVIADO A OTRA BASE DE DATOS

DO GAD4S

? TSTRING(SECONDS()-T)

#INCLUDE "ALGOS.PRG"
#INCLUDE "ALGOC.PRG"
```

P 9 (PROGRAMA 9 : ESTIMADORES CON UNA MUESTRA ALEATORIA ESTRATIFICADA POR DISTRITO ELECTORAL, CALCULANDO CUATRO ESTIMADORES Y GENERANDO GRUPOS ALEATORIOS INDEPENDIENTES Y DEPENDIENTES CON K=2,4)

```
**********************************
*P9.PRG
TIEM=SECONDS()
SET COLOR TO W/B,B/W,B
BANDERA="V"
BVEZ=1
BVEQ=1
CLEAR
@ 2,10 SAY " ESTRATIFICADO POR DISTRITO"
@ 5,10 SAY " CUANTAS VECES "
@ 5,30 GET BVEQ PICTURE "9999"
READ
DO WHILE BVEZ<=BVEQ
CASO="SIN"
CLS
SET SAFETY OFF
SET DECIMALS TO 10
SET EXACT ON
USE VALIDO3
ZAP
USE
USE VALIDO
TER="NO"
ZAP
E=1
EC=LTRIM(STR(E))
MB=426
DO WHILE E<=13  && PARA CADA ESTRATO :
H=0
EC=LTRIM(STR(E))
USE ED&EC
```

AVERAGE VALIDOS TO MB&EC && $MB\&EC = \overline{M}_h = \dfrac{\sum_{i=1}^{N_h} M_{hi}}{N_h}$

N=RECCOUNT() && Nh = N

M=((125)*N)/2081 && $n_h = n\dfrac{N_h}{N}$ asignación proporcional

M=ROUND(M,0)

? " ESTRATO "+LTRIM(STR(E))+" numero "+LTRIM(STR(BVEZ))+" de "+LTRIM(STR(BVEQ))

? "Muestra de "+ltrim(STR(M))+" elementos ";

+"&CASO"+" reemplazamiento de una poblacien de tamaoo "+LTRIM(STR(N))

?

? BANDERA

SUM VALIDOS TO STOTAL&EC

MUES(N,M,"ED&EC")

**muestra del estrato

IF TER="NO"

CLEAR

USE VALIDO && usa la muestra

LIST NUMERO

?

SET DECIMAL TO 10

SUM PRI TO SPRI

SUM PAN TO SPAN

SUM PRD TO SPRD

SUM OTROS TO SOTROS

SUM VALIDOS TO TTOTAL && $TTOTAL = \sum_{i=1}^{n_h} M_{hi}$ para el estrato h

AVERAGE PRI TO YBPRI

YBPRI=YBPRI/MB && Parte del estimador insesgado $\dfrac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h\overline{M}} = \dfrac{\sum_{i=1}^{n_h} y_{hi}}{n_h\overline{M}}$

AVERAGE PAN TO YBPAN

YBPAN=YBPAN/MB

AVERAGE PRD TO YBPRD

YBPRD=YBPRD/MB

AVERAGE OTROS TO YBOTROS

YBOTROS=YBOTROS/MB

AVERAGE VALIDOS TO PTOTAL

AVERAGE PRI/VALIDOS TO E2PRI && Parte del estimador media de medias $\dfrac{\sum_{i=1}^{n_h} \bar{y}_{hi\cdot}}{n_h}$

```
AVERAGE PAN/VALIDOS  TO E2PAN
AVERAGE PRD/VALIDOS  TO E2PRD
AVERAGE OTROS/VALIDOS  TO E2OTROS


AVERAGE PRI/MB&EC TO E3PRI    && Parte del estimador insesgado por estrato
```

$$\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{n_h\overline{M}_h} = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h\overline{M}_h}$$

```
AVERAGE PAN/MB&EC TO E3PAN
AVERAGE PRD/MB&EC TO E3PRD
AVERAGE OTROS/MB&EC TO E3OTROS


E4PRI=SPRI/TTOTAL && Parte del estimador de razón por estrato
```

$$\frac{\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h} M_{hi}} = \frac{\sum_{i=1}^{n_h} y_{hi}}{\sum_{i=1}^{n_h} M_{hi}}$$

```
E4PAN=SPAN/TTOTAL
E4OTROS=SOTROS/TTOTAL
E4PRD=SPRD/TTOTAL
ENDIF
?
USE VALIDO3
APPEND BLANK
REPLACE CPRI WITH YBPRI,CPAN WITH YBPAN      && EST. 1
REPLACE CPRD WITH YBPRD  &&,CVAL WITH YBVAL
REPLACE CMU WITH M,COTROS WITH YBOTROS
******
REPLACE CPRI2 WITH E2PRI           && EST. 2
REPLACE CPRI3 WITH E3PRI
REPLACE CPRI4 WITH E4PRI
REPLACE CPAN2 WITH E2PAN
REPLACE CPAN3 WITH E3PAN
REPLACE CPAN4 WITH E4PAN
REPLACE CPRD2 WITH E2PRD
REPLACE CPRD3 WITH E3PRD
REPLACE CPRD4 WITH E4PRD
REPLACE COTROS2 WITH E2OTROS
REPLACE COTROS3 WITH E3OTROS
REPLACE COTROS4 WITH E4OTROS

****DATOS DEL ESTRATO:
REPLACE CNH WITH N   && Nh
REPLACE CMH WITH STOTAL&EC  && Mh
USE
EC=EC+".DBF"
!COPY VALIDO.DBF MUED&EC
USE VALIDO

ZAP
```

```
E=E+1   && OTRO ESTRATO

ENDDO
USE VALIDO3
SUM CMU TO SMU
SET DECIMALS TO 10
GO TOP
MB=426
MB2=MB**2
Q=2081*MB && N*MBARRA
?
SUM CPRI*CNH TO SUPRI
SUM CPRI2*CNH TO SUPRI2
SUM CPRI3*CNH TO SUPRI3
SUM CPRI4*CNH TO SUPRI4
SUM CPAN*CNH TO SUPAN
SUM CPAN2*CNH TO SUPAN2
SUM CPAN3*CNH TO SUPAN3
SUM CPAN4*CNH TO SUPAN4
SUM CPRD*CNH TO SUPRD
SUM CPRD2*CNH TO SUPRD2
SUM CPRD3*CNH TO SUPRD3
SUM CPRD4*CNH TO SUPRD4
SUM COTROS*CNH TO SUOTROS
SUM COTROS2*CNH TO SUOTROS2
SUM COTROS3*CNH TO SUOTROS3
SUM COTROS4*CNH TO SUOTROS4
```

MPRI=SUPRI/2081 && Media del estimador insegado $\hat{\bar{Y}}_{st} = \frac{1}{N\bar{M}}\sum_{h=1}^{L} N_h \frac{1}{n_h}\sum_{i=1}^{n_h} M_{hi}\bar{y}_{hi\cdot}$

MPRI2=SUPRI2/2081 && Media del estimador media de medias por estrato $\hat{\bar{Y}}_{mst} = \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}\bar{y}_{hi\cdot}}{n_h}$

MPRI3=SUPRI3/2081 && Media del estimador insesgado por estrato $\hat{\bar{Y}}_{ust} = \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{n_h\bar{M}_h}$

MPRI4=SUPRI4/2081 && Media del estimador de razón por estrato $\hat{\bar{Y}}_{rst} = \sum_{h=1}^{L}\frac{N_h}{N}\frac{\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi\cdot}}{\sum_{i=1}^{n_h}M_{hi}}$

```
MPAN=SUPAN/2081
MPAN2=SUPAN2/2081
MPAN3=SUPAN3/2081
MPAN4=SUPAN4/2081

MPRD=SUPRD/2081
MPRD2=SUPRD2/2081
MPRD3=SUPRD3/2081
```

```
MPRD4=SUPRD4/2081
MOTROS=SUOTROS/2081
MOTROS2=SUOTROS2/2081
MOTROS3=SUOTROS3/2081
MOTROS4=SUOTROS4/2081

USE RESULMED
APPEND BLANK
REPLACE RPRI WITH MPRI,RPAN WITH MPAN
REPLACE RPRI2 WITH MPRI2,RPRI3 WITH MPRI3
REPLACE RPRI4 WITH MPRI4
REPLACE RPAN2 WITH MPAN2,RPAN3 WITH MPAN3
REPLACE RPAN4 WITH MPAN4
**
REPLACE RPRD WITH MPRD,RPRD2 WITH MPRD2
REPLACE RPRD4 WITH MPRD4
REPLACE RPRD3 WITH MPRD3,ROTROS WITH MOTROS
REPLACE ROTROS2 WITH MOTROS2,ROTROS3 WITH MOTROS3
REPLACE ROTROS4 WITH MOTROS4

SET DECIMALS TO 4

BVEZ=BVEZ+1
BANDERA="R"
? BANDERA
DO P10
DO P11
DO P12
DO P13
ENDDO
**********RETURN
? TSTRING(SECONDS()-TIEM)
? CHR(7)
#INCLUDE "ALGOC.PRG"
#INCLUDE "ALGOS.PRG"
```

P 10 (PROGRAMA 10 : GRUPOS ALEATORIOS INDEPENDIENTES CON K=2 DE UNA MUESTRA ALEATORIA ESTRATIFICADA POR DISTRITO ELECTORAL)

```
**************************
*P10.PRG
USE HBO
ZAP
USE

SET COLOR TO W/B,B/W,B
BANDERA="V"
CVEZ=1
CVEQ=2
CLEAR
DO WHILE CVEZ<=CVEQ
CASO="SIN"
CLS
```

```
SET SAFETY OFF
SET DECIMALS TO 10
SET EXACT ON
USE VALIDO3
ZAP
USE
USE VALIDO
TER="NO"
ZAP
E=1
EC=LTRIM(STR(E))
MB=426
DO WHILE E<=13
H=0
EC=LTRIM(STR(E))
USE ED&EC
AVERAGE VALIDOS TO MB&EC
 N=RECCOUNT() && Nh = N
USE MUED&EC
N2=RECCOUNT()
M=2
M=ROUND((1/M)*N2,0)
 M=ROUND(M,0)
? " ESTRATO  "+LTRIM(STR(E))+" numero  "+LTRIM(STR(CVEZ))+" de "+LTRIM(STR(CVEQ))
? "Muestra de "+ltrim(STR(M))+" elementos ";
+"&CASO"+" reemplazamiento de una poblacien de tamaño "+LTRIM(STR(N))
?
? BANDERA
SUM VALIDOS TO STOTAL&EC

MUEC(N2,M,"MUED&EC")
***************************************************MUESTRA
IF TER="NO"
CLEAR
USE VALIDO    && USA LA MUESTRA
?
SET DECIMAL TO 10
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
SUM VALIDOS TO TTOTAL


AVERAGE PRI/VALIDOS TO E2PRI   && EST. 2
AVERAGE PAN/VALIDOS TO E2PAN
AVERAGE PRD/VALIDOS TO E2PRD
AVERAGE OTROS/VALIDOS TO E2OTROS


E4PRI=SPRI/TTOTAL && ESTIMADOR 4
E4PAN=SPAN/TTOTAL
```

```
E4OTROS=SOTROS/TTOTAL
E4PRD=SPRD/TTOTAL
ENDIF
USE VALIDO3
APPEND BLANK
REPLACE CMU WITH M &&.COTROS WITH YBOTROS
REPLACE CPRI2 WITH E2PRI        && EST. 2
REPLACE CPRI4 WITH E4PRI
REPLACE CPAN2 WITH E2PAN
REPLACE CPAN4 WITH E4PAN
REPLACE CPRD2 WITH E2PRD
REPLACE CPRD4 WITH E4PRD
REPLACE COTROS2 WITH E2OTROS
REPLACE COTROS4 WITH E4OTROS

****DATOS DEL ESTRATO:
REPLACE CNH WITH N  && Nh
REPLACE CMH WITH STOTAL&EC  && Mh
USE
EC=EC+".DBF"
USE VALIDO
ZAP

E=E+1  && OTRO ESTRATO

ENDDO
USE VALIDO3
SUM CMU TO SMU
SET DECIMALS TO 10
GO TOP
MB=426   && PARA ESTRAT URB/RUR 622478 Y PARA DISTRITO 69164
MB2=MB**2
Q=2081*MB && N*MBARRA

SUM CPRI2*CNH TO SUPRI2  && "                EST. 2
SUM CPRI4*CNH TO SUPRI4
SUM CPAN2*CNH TO SUPAN2
SUM CPAN4*CNH TO SUPAN4
SUM CPRD2*CNH TO SUPRD2
SUM CPRD4*CNH TO SUPRD4
SUM COTROS2*CNH TO SUOTROS2
SUM COTROS4*CNH TO SUOTROS4

MPRI2=SUPRI2/2081 && MEDIA ESTRA. EST 2
MPRI4=SUPRI4/2081
MPAN2=SUPAN2/2081
MPAN4=SUPAN4/2081

MPRD2=SUPRD2/2081
MPRD4=SUPRD4/2081
MOTROS2=SUOTROS2/2081
MOTROS4=SUOTROS4/2081
```

```
USE HBO
APPEND BLANK
REPLACE RPRI2 WITH MPRI2
REPLACE RPRI4 WITH MPRI4
REPLACE RPAN2 WITH MPAN2
REPLACE RPAN4 WITH MPAN4
**
REPLACE RPRD2 WITH MPRD2
REPLACE RPRD4 WITH MPRD4
REPLACE ROTROS2 WITH MOTROS2
REPLACE ROTROS4 WITH MOTROS4

SET DECIMALS TO 10   && TENUIA 4

CVEZ=CVEZ+1
BANDERA="R"

ENDDO
USE HBO
AVERAGE RPRI2 TO P2
AVERAGE RPRI4 TO P4
AVERAGE RPAN2 TO R2
AVERAGE RPAN4 TO R4
AVERAGE RPRD2 TO Q2
AVERAGE RPRD4 TO Q4


SUM (RPRI2-P2)**2 TO SC2P1
SUM (RPRI4-P4)**2 TO SC4P1

VAG2P1=SC2P1/2
VAG4P1=SC4P1/2
SUM (RPAN2-R2)**2 TO SC2P2
SUM (RPAN4-R4)**2 TO SC4P2
VAG2P2=SC2P2/2
VAG4P2=SC4P2/2

SUM (RPRD2-Q2)**2 TO SC2P3
SUM (RPRD4-Q4)**2 TO SC4P3
VAG2P3=SC2P3/2
VAG4P3=SC4P3/2

USE VAGED
APPEND BLANK
REPLACE CMEG WITH P2   && MEDIA DEL P. 1 PARA EL EST. 2
REPLACE CVAG WITH VAG2P1  && VAR. DEL EST.2 PART. 1
REPLACE CMEG2 WITH P4   && MEDIA DEL P. 2 PARA EL EST. 4
REPLACE CVAG2 WITH VAG4P1

REPLACE CMEG5 WITH R2
REPLACE CVAG5 WITH VAG2P2
```

```
REPLACE CMEG6 WITH R4
REPLACE CVAG6 WITH VAG4P2

REPLACE CMEG9 WITH Q2
REPLACE CVAG9 WITH VAG2P3
REPLACE CMEG10 WITH Q4
REPLACE CVAG10 WITH VAG4P3
```

P 11 (PROGRAMA 11 : GRUPOS ALEATORIOS INDEPENDIENTES CON K=4 DE UNA MUESTRA ESTRATIFICADA POR DISTRITO ELECTORAL)

```
*P11.PRG
**
USE HBO
ZAP
USE
SET COLOR TO W/B,B/W,B
BANDERA="V"
CVEZ=1
CVEQ=4
CLEAR
DO WHILE CVEZ<=CVEQ
CASO="SIN"
CLS
SET SAFETY OFF
SET DECIMALS TO 10
SET EXACT ON
USE VALIDO3
ZAP
USE
USE VALIDO
TER="NO"
ZAP
E=1
EC=LTRIM(STR(E))
MB=426
DO WHILE E<=13
H=0
EC=LTRIM(STR(E))
USE ED&EC
AVERAGE VALIDOS TO MB&EC
 N=RECCOUNT() && Nh = N
USE MUED&EC
N2=RECCOUNT()
M=4    &&
M=ROUND((1/M)*N2,0)
 M=ROUND(M,0)
? " ESTRATO  "+LTRIM(STR(E))+" numero  "+LTRIM(STR(CVEZ))+" de "+LTRIM(STR(CVEQ))
? "Muestra de "+ltrim(STR(M))+" elementos ";
+"&CASO"+" reemplazamiento de una poblaci¢n de tama¤o "+LTRIM(STR(N))
?
? BANDERA
```

```
*WAIT

*? CHR(7)
 SUM VALIDOS TO STOTAL&EC

MUEC(N2,M,"MUED&EC")
**********************************************MUESTRA
IF TER="NO"
CLEAR
USE VALIDO     && USA LA MUESTRA
?
SET DECIMAL TO 10
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
SUM VALIDOS TO TTOTAL

AVERAGE PRI/VALIDOS TO E2PRI    && EST. 2
AVERAGE PAN/VALIDOS TO E2PAN
AVERAGE PRD/VALIDOS TO E2PRD
AVERAGE OTROS/VALIDOS TO E2OTROS


E4PRI=SPRI/TTOTAL && ESTIMADOR 4
E4PAN=SPAN/TTOTAL
E4OTROS=SOTROS/TTOTAL
E4PRD=SPRD/TTOTAL

ENDIF
?
** COPIA A VALIDO3 Y POR CADA ESTRATO LOS VALORES:
USE VALIDO3
APPEND BLANK
REPLACE CMU WITH M &&,COTROS WITH YBOTROS
******
REPLACE CPRI2 WITH E2PRI          && EST. 2
REPLACE CPRI4 WITH E4PRI
REPLACE CPAN2 WITH E2PAN
REPLACE CPAN4 WITH E4PAN
REPLACE CPRD2 WITH E2PRD
REPLACE CPRD4 WITH E4PRD
REPLACE COTROS2 WITH E2OTROS
REPLACE COTROS4 WITH E4OTROS

****DATOS DEL ESTRATO:
REPLACE CNH WITH N  && Nh
REPLACE CMH WITH STOTAL&EC  && Mh
USE
EC=EC+".DBF"
*!COPY VALIDO.DBF MUED&EC
USE VALIDO
```

```
ZAP

E=E+1   && OTRO ESTRATO

ENDDO
USE VALIDO3
SUM CMU TO SMU
SET DECIMALS TO 10
GO TOP
MB=426   && PARA ESTRAT URB/RUR 622478 Y PARA DISTRITO 69164
MB2=MB**2
Q=2081*MB && N*MBARRA
?
SUM CPRI2*CNH TO SUPRI2   && "                    EST. 2
SUM CPRI4*CNH TO SUPRI4
SUM CPAN2*CNH TO SUPAN2
SUM CPAN4*CNH TO SUPAN4
SUM CPRD2*CNH TO SUPRD2
SUM CPRD4*CNH TO SUPRD4
SUM COTROS2*CNH TO SUOTROS2
SUM COTROS4*CNH TO SUOTROS4

MPRI2=SUPRI2/2081 && MEDIA ESTRA. EST 2
MPRI4=SUPRI4/2081
MPAN2=SUPAN2/2081
MPAN4=SUPAN4/2081

MPRD2=SUPRD2/2081
MPRD4=SUPRD4/2081
MOTROS2=SUOTROS2/2081
MOTROS4=SUOTROS4/2081
USE HBO
APPEND BLANK
REPLACE RPRI2 WITH MPRI2
REPLACE RPRI4 WITH MPRI4
REPLACE RPAN2 WITH MPAN2
REPLACE RPAN4 WITH MPAN4
**
REPLACE RPRD2 WITH MPRD2
REPLACE RPRD4 WITH MPRD4
REPLACE ROTROS2 WITH MOTROS2
REPLACE ROTROS4 WITH MOTROS4

SET DECIMALS TO 4

CVEZ=CVEZ+1
BANDERA="R"
ENDDO
**********RETURN
USE HBO
? RECCOUNT()
```

```
AVERAGE RPRI2 TO P2
AVERAGE RPRI4 TO P4
AVERAGE RPAN2 TO R2
AVERAGE RPAN4 TO R4

AVERAGE RPRD2 TO Q2
AVERAGE RPRD4 TO Q4

SUM (RPRI2-P2)**2 TO SC2P1
SUM (RPRI4-P4)**2 TO SC4P1
VAG2P1=SC2P1/12
VAG4P1=SC4P1/12
SUM (RPAN2-R2)**2 TO SC2P2
SUM (RPAN4-R4)**2 TO SC4P2
VAG2P2=SC2P2/12
VAG4P2=SC4P2/12
SUM (RPRD2-Q2)**2 TO SC2P3
SUM (RPRD4-Q4)**2 TO SC4P3
VAG2P3=SC2P3/12
VAG4P3=SC4P3/12

USE VAGE2D
APPEND BLANK
REPLACE CMEG WITH P2
REPLACE CVAG WITH VAG2P1
REPLACE CMEG2 WITH P4
REPLACE CVAG2 WITH VAG4P1

REPLACE CMEG5 WITH R2
REPLACE CVAG5 WITH VAG2P2
REPLACE CMEG6 WITH R4
REPLACE CVAG6 WITH VAG4P2

REPLACE CMEG9 WITH Q2
REPLACE CVAG9 WITH VAG2P3
REPLACE CMEG10 WITH Q4
REPLACE CVAG10 WITH VAG4P3
```

P 12 (PROGRAMA 12 : GRUPOS ALEATORIOS DEPENDIENTES CON K=2 DE UNA MUESTRA ALEATORIA ESTRATIFICADA POR DISTRITO ELECTORAL)

```
*P12.PRG
USE HBO
ZAP
USE
SET COLOR TO W/B,B/W,B
BANDERA="V"
CVEZ=1
CVEQ=2
CLEAR
DO WHILE CVEZ<=CVEQ
CASO="SIN"
```

```
CLS
SET SAFETY OFF
SET DECIMALS TO 10
SET EXACT ON
USE VALIDO3
ZAP
USE
USE VALIDO
TER="NO"
ZAP
E=1
EC=LTRIM(STR(E))
MB=426
DO WHILE E<=13
H=0
EC=LTRIM(STR(E))
USE ED&EC
AVERAGE VALIDOS TO MB&EC
 N=RECCOUNT() && Nh = N
USE MUED&EC
N2=RECCOUNT()
? " ESTRATO  "+LTRIM(STR(E))+" numero  "+LTRIM(STR(CVEZ))+" de "+LTRIM(STR(CVEQ))
*? "Muestra de "+ltrim(STR(M))+" elementos ";
*+"&CASO"+" reemplazamiento de una población de tamaño "+LTRIM(STR(N))
?
? BANDERA
SUM VALIDOS TO STOTAL&EC
***********************
DVEZ=CVEZ
USE VALIDO
ZAP
IF CVEZ=2
 DVEZ=0
ENDIF
APPEND FROM MUED&EC FOR MOD(RECNO(),2)=DVEZ

**********************************************MUESTRA
IF TER="NO"
CLEAR
USE VALIDO     && USA LA MUESTRA DEL ESTRATO
?
SET DECIMAL TO 10
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
SUM VALIDOS TO TTOTAL


AVERAGE PRI/VALIDOS TO E2PRI   && EST. 2
AVERAGE PAN/VALIDOS TO E2PAN
```

```
AVERAGE PRD/VALIDOS TO E2PRD
AVERAGE OTROS/VALIDOS TO E2OTROS


E4PRI=SPRI/TTOTAL && ESTIMADOR 4
E4PAN=SPAN/TTOTAL
E4OTROS=SOTROS/TTOTAL
E4PRD=SPRD/TTOTAL

ENDIF
?
** COPIA A VALIDO3 Y POR CADA ESTRATO LOS VALORES:
USE VALIDO3
APPEND BLANK
******
REPLACE CPRI2 WITH E2PRI        && EST. 2
REPLACE CPRI4 WITH E4PRI
REPLACE CPAN2 WITH E2PAN
REPLACE CPAN4 WITH E4PAN
REPLACE CPRD2 WITH E2PRD
REPLACE CPRD4 WITH E4PRD
REPLACE COTROS2 WITH E2OTROS
REPLACE COTROS4 WITH E4OTROS

****DATOS DEL ESTRATO:
REPLACE CNH WITH N  && Nh
REPLACE CMH WITH STOTAL&EC  && Mh
USE
EC=EC+".DBF"
*!COPY VALIDO.DBF MUED&EC
USE VALIDO
ZAP

E=E+1  && OTRO ESTRATO

ENDDO
USE VALIDO3
SUM CMU TO SMU
SET DECIMALS TO 10
GO TOP
MB=426   && PARA ESTRAT URB/RUR 622478 Y PARA DISTRITO 69164
MB2=MB**2
Q=2081*MB && N*MBARRA
?
SUM CPRI2*CNH TO SUPRI2  && "              EST. 2
SUM CPRI4*CNH TO SUPRI4
SUM CPAN2*CNH TO SUPAN2
SUM CPAN4*CNH TO SUPAN4
SUM CPRD2*CNH TO SUPRD2
SUM CPRD4*CNH TO SUPRD4
SUM COTROS2*CNH TO SUOTROS2
```

```
SUM COTROS4*CNH TO SUOTROS4

MPRI2=SUPRI2/2081 && MEDIA ESTRA. EST 2
MPRI4=SUPRI4/2081
MPAN2=SUPAN2/2081
MPAN4=SUPAN4/2081

MPRD2=SUPRD2/2081
MPRD4=SUPRD4/2081
MOTROS2=SUOTROS2/2081
MOTROS4=SUOTROS4/2081
*

USE HBO
APPEND BLANK
REPLACE RPRI2 WITH MPRI2
REPLACE RPRI4 WITH MPRI4
REPLACE RPAN2 WITH MPAN2
REPLACE RPAN4 WITH MPAN4
**
REPLACE RPRD2 WITH MPRD2
REPLACE RPRD4 WITH MPRD4
REPLACE ROTROS2 WITH MOTROS2
REPLACE ROTROS4 WITH MOTROS4

SET DECIMALS TO 10   && TENUIA 4

CVEZ=CVEZ+1
BANDERA="R"

ENDDO
**********RETURN
USE HBO
? RECCOUNT()

AVERAGE RPRI2 TO P2
AVERAGE RPRI4 TO P4
AVERAGE RPAN2 TO R2
AVERAGE RPAN4 TO R4

AVERAGE RPRD2 TO Q2
AVERAGE RPRD4 TO Q4

SUM (RPRI2-P2)**2 TO SC2P1
SUM (RPRI4-P4)**2 TO SC4P1
VAG2P1=SC2P1/2
VAG4P1=SC4P1/2
SUM (RPAN2-R2)**2 TO SC2P2
SUM (RPAN4-R4)**2 TO SC4P2
VAG2P2=SC2P2/2
VAG4P2=SC4P2/2
SUM (RPRD2-Q2)**2 TO SC2P3
```

```
SUM (RPRD4-Q4)**2 TO SC4P3
VAG2P3=SC2P3/2
VAG4P3=SC4P3/2

USE VAGED
GO BOTTOM
REPLACE CMEG3 WITH P2   && MEDIA DEL P. 1 PARA EL EST. 2
REPLACE CVAG3 WITH VAG2P1  && VAR. DEL EST.2 PART. 1
REPLACE CMEG4 WITH P4   && MEDIA DEL P. 2 PARA EL EST. 4
REPLACE CVAG4 WITH VAG4P1

REPLACE CMEG7 WITH R2
REPLACE CVAG7 WITH VAG2P2
REPLACE CMEG8 WITH R4
REPLACE CVAG8 WITH VAG4P2


REPLACE CMEG11 WITH Q2
REPLACE CVAG11 WITH VAG2P3
REPLACE CMEG12 WITH Q4
REPLACE CVAG12 WITH VAG4P3
```

P 13 (PROGRAMA 13 : GRUPOS ALEATORIOS DEPENDIENTES CON K=4 DE UNA MUESTRA ALEATORIA ESTRATIFICADA POR DISTRITO ELECTORAL)

```
*P13.PRG
**
USE HBO
ZAP
USE

SET COLOR TO W/B,B/W,B
BANDERA="V"
CVEZ=1
CVEQ=4
CLEAR
DO WHILE CVEZ<=CVEQ
CASO="SIN"
CLS
SET SAFETY OFF
SET DECIMALS TO 10
SET EXACT ON
USE VALIDO3
ZAP
USE
USE VALIDO
TER="NO"
ZAP
E=1
EC=LTRIM(STR(E))
MB=426
DO WHILE E<=13
```

```
H=0
EC=LTRIM(STR(E))
USE ED&EC
AVERAGE VALIDOS TO MB&EC
 N=RECCOUNT() && Nh = N
? N
USE MUED&EC
N2=RECCOUNT()
? " ESTRATO  "+LTRIM(STR(E))+" numero  "+LTRIM(STR(CVEZ))+" de "+LTRIM(STR(CVEQ))
*? "Muestra de "+ltrim(STR(M))+" elementos ";
*+"&CASO"+" reemplazamiento de una poblacien de tamaño "+LTRIM(STR(N))
?
? BANDERA

SUM VALIDOS TO STOTAL&EC
************************
DVEZ=CVEZ
USE VALIDO
ZAP
IF CVEZ=4
 DVEZ=0
ENDIF
APPEND FROM MUED&EC FOR MOD(RECNO(),4)=DVEZ



**************************************************MUESTRA
IF TER="NO"
CLEAR
USE VALIDO     && USA LA MUESTRA DEL ESTRATO
?
SET DECIMAL TO 10
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
SUM VALIDOS TO TTOTAL


AVERAGE PRI/VALIDOS TO E2PRI    && EST. 2
AVERAGE PAN/VALIDOS TO E2PAN
AVERAGE PRD/VALIDOS TO E2PRD
AVERAGE OTROS/VALIDOS TO E2OTROS


E4PRI=SPRI/TTOTAL && ESTIMADOR 4
E4PAN=SPAN/TTOTAL
E4OTROS=SOTROS/TTOTAL
E4PRD=SPRD/TTOTAL
ENDIF
?
** COPIA A VALIDO3 Y POR CADA ESTRATO LOS VALORES:
```

```
USE VALIDO3
APPEND BLANK
******
REPLACE CPRI2 WITH E2PRI        && EST. 2
REPLACE CPRI4 WITH E4PRI
REPLACE CPAN2 WITH E2PAN
REPLACE CPAN4 WITH E4PAN
REPLACE CPRD2 WITH E2PRD
REPLACE CPRD4 WITH E4PRD
REPLACE COTROS2 WITH E2OTROS
REPLACE COTROS4 WITH E4OTROS

****DATOS DEL ESTRATO:
REPLACE CNH WITH N  && Nh
REPLACE CMH WITH STOTAL&EC  && Mh
USE
EC=EC+".DBF"
USE VALIDO
ZAP

E=E+1  && OTRO ESTRATO

ENDDO
USE VALIDO3
SUM CMU TO SMU
SET DECIMALS TO 10
GO TOP
MB=426   && PARA ESTRAT URB/RUR 622478 Y PARA DISTRITO 69164
MB2=MB**2
Q=2081*MB && N*MBARRA
?
SUM CPRI2*CNH TO SUPRI2  && "                EST. 2
SUM CPRI4*CNH TO SUPRI4
SUM CPAN2*CNH TO SUPAN2
SUM CPAN4*CNH TO SUPAN4
SUM CPRD2*CNH TO SUPRD2
SUM CPRD4*CNH TO SUPRD4
SUM COTROS2*CNH TO SUOTROS2
SUM COTROS4*CNH TO SUOTROS4

MPRI2=SUPRI2/2081 && MEDIA ESTRA. EST 2
MPRI4=SUPRI4/2081
MPAN2=SUPAN2/2081
MPAN4=SUPAN4/2081
MPRD2=SUPRD2/2081
MPRD4=SUPRD4/2081
MOTROS2=SUOTROS2/2081
MOTROS4=SUOTROS4/2081
*

USE HBO
APPEND BLANK
```

```
REPLACE RPRI2 WITH MPRI2
REPLACE RPRI4 WITH MPRI4
REPLACE RPAN2 WITH MPAN2
REPLACE RPAN4 WITH MPAN4
**
REPLACE RPRD2 WITH MPRD2
REPLACE RPRD4 WITH MPRD4
REPLACE ROTROS2 WITH MOTROS2
REPLACE ROTROS4 WITH MOTROS4

SET DECIMALS TO 10   && TENIJIA 4

CVEZ=CVEZ+1
BANDERA="R"

ENDDO
**********RETURN
USE HBO
? RECCOUNT()

AVERAGE RPRI2 TO P2
AVERAGE RPRI4 TO P4
AVERAGE RPAN2 TO R2
AVERAGE RPAN4 TO R4
AVERAGE RPRD2 TO Q2
AVERAGE RPRD4 TO Q4

SUM (RPRI2-P2)**2 TO SC2P1
SUM (RPRI4-P4)**2 TO SC4P1
VAG2P1=SC2P1/12
VAG4P1=SC4P1/12
SUM (RPAN2-R2)**2 TO SC2P2
SUM (RPAN4-R4)**2 TO SC4P2
VAG2P2=SC2P2/12
VAG4P2=SC4P2/12
SUM (RPRD2-Q2)**2 TO SC2P3
SUM (RPRD4-Q4)**2 TO SC4P3
VAG2P3=SC2P3/12
VAG4P3=SC4P3/12

USE VAGE2D

GO BOTTOM
REPLACE CMEG3 WITH P2   && MEDIÁ DEL P. 1 PARA EL EST. 2
REPLACE CVAG3 WITH VAG2P1  && VAR. DEL EST.2 PART. 1
REPLACE CMEG4 WITH P4   && MEDIA DEL P. 2 PARA EL EST..4
REPLACE CVAG4 WITH VAG4P1

REPLACE CMEG7 WITH R2
REPLACE CVAG7 WITH VAG2P2
REPLACE CMEG8 WITH R4
REPLACE CVAG8 WITH VAG4P2
```

```
REPLACE CMEG11 WITH Q2
REPLACE CVAG11 WITH VAG2P3
REPLACE CMEG12 WITH Q4
REPLACE CVAG12  WITH VAG4P3
```

P 14 (PROGRAMA 14 : ESTIMADORES CON UNA MUESTRA ALEATORIA ESTRATIFICADA Y SISTEMÁTICA CON ARRANQUE ALEATORIO INDEPENDIENTE EN CADA ESTRATO EN UN MUESTREO ESTRATIFICADO URBANO/RURAL POR CASILLAS PARA EL ESTADO DE GUANAJUATO)

```
****************************
*P14.PRG
TIEM=SECONDS()
SET COLOR TO W/B,B/W,B
BANDERA="V"
BVEZ=1
BVEQ=1
CLEAR
@ 2,10 SAY " ESTRATIFICADO POR DISTRITO"
@ 5,10 SAY " CUANTAS VECES "
@ 5,30 GET BVEQ PICTURE "9999"
READ
DO WHILE BVEZ<=BVEQ
CASO="SIN"
CLS
SET SAFETY OFF
SET DECIMALS TO 10
SET EXACT ON
USE VALIDO3
ZAP
USE
USE VALIDO
TER="NO"
ZAP
E=1
EC=LTRIM(STR(E))
MB=283
DO WHILE E<=2
H=0
EC=LTRIM(STR(E))
USE EUR&EC
AVERAGE VALIDOS TO MB&EC
 N=RECCOUNT() && Nh = N
 M=((176)*N)/4399

 M=ROUND(M,0)
? " ESTRATO  "+LTRIM(STR(E))+" numero  "+LTRIM(STR(BVEZ))+" de "+LTRIM(STR(BVEQ))
? "Muestra de "+ltrim(STR(M))+" elementos ";
+"&CASO"+" reemplazamiento de una poblacien de tamaoo "+LTRIM(STR(N))
?
```

```
? BANDERA
*WAIT

SUM VALIDOS TO STOTAL&EC

MUES(25,1,"EUR&EC")
*************************************************MUESTRA
IF TER="NO"
CLEAR
USE VALIDO    && USA LA MUESTRA DEL ESTRATO
***AQUI SISTEMATICO
  STORE NUMERO TO NUM
  NUMER=NUM
  P=25
  W=1
  USE VALIDO2
  ZAP
  DO WHILE W<=M
    APPEND RECORD NUM FROM EUR&EC
    NUM=NUM+P
    IF NUM>=N
      NUM=1
    ENDIF
  W=W+1
  ENDDO



SET DECIMAL TO 10
SUM PRI TO SPRI
SUM PAN TO SPAN
SUM PRD TO SPRD
SUM OTROS TO SOTROS
SUM VALIDOS TO TTOTAL


AVERAGE PRI TO YBPRI    && EST. 1
YBPRI=YBPRI/MB
AVERAGE PAN TO YBPAN
YBPAN=YBPAN/MB
AVERAGE PRD TO YBPRD
YBPRD=YBPRD/MB
AVERAGE OTROS TO YBOTROS
YBOTROS=YBOTROS/MB
AVERAGE VALIDOS TO PTOTAL


AVERAGE PRI/VALIDOS TO E2PRI    && EST. 2
AVERAGE PAN/VALIDOS TO E2PAN
AVERAGE PRD/VALIDOS TO E2PRD
AVERAGE OTROS/VALIDOS TO E2OTROS
```

```
AVERAGE PRI/MB&EC TO E3PRI    && EST. 3
AVERAGE PAN/MB&EC TO E3PAN
AVERAGE PRD/MB&EC TO E3PRD
AVERAGE OTROS/MB&EC TO E3OTROS


E4PRI=SPRI/TTOTAL && ESTIMADOR 4
E4PAN=SPAN/TTOTAL
E4OTROS=SOTROS/TTOTAL
E4PRD=SPRD/TTOTAL
ENDIF
?
** COPIA A VALIDO3 Y POR CADA ESTRATO LOS VALORES:
USE VALIDO3
APPEND BLANK
REPLACE CPRI WITH YBPRI,CPAN WITH YBPAN     && EST. 1
REPLACE CPRD WITH YBPRD  &&,CVAL WITH YBVAL
REPLACE CMU WITH M,COTROS WITH YBOTROS
******
REPLACE CPRI2 WITH E2PRI          && EST. 2
REPLACE CPRI3 WITH E3PRI
REPLACE CPRI4 WITH E4PRI
REPLACE CPAN2 WITH E2PAN
REPLACE CPAN3 WITH E3PAN
REPLACE CPAN4 WITH E4PAN
REPLACE CPRD2 WITH E2PRD
REPLACE CPRD3 WITH E3PRD
REPLACE CPRD4 WITH E4PRD
REPLACE COTROS2 WITH E2OTROS
REPLACE COTROS3 WITH E3OTROS
REPLACE COTROS4 WITH E4OTROS

****DATOS DEL ESTRATO:
REPLACE CNH WITH N  && Nh
REPLACE CMH WITH STOTAL&EC  && Mh
USE
EC=EC+".DBF"
!COPY VALIDO2.DBF MUEUR&EC
USE VALIDO2
ZAP

E=E+1   && OTRO ESTRATO
ENDDO
USE VALIDO3
*LIST MUESTRA
SUM CMU TO SMU
SET DECIMALS TO 10
GO TOP
MB=283    && PARA ESTRAT URB/RUR 622478 Y PARA DISTRITO 69164
MB2=MB**2
```

```
Q=4399*MB && N*MBARRA
?
SUM CPRI*CNH TO SUPRI    && PRIMERA PARTE DE LA MEDIA EST .1
SUM CPRI2*CNH TO SUPRI2  && "                    EST. 2
SUM CPRI3*CNH TO SUPRI3  &&       "              EST. 3
SUM CPRI4*CNH TO SUPRI4
SUM CPAN*CNH TO SUPAN
SUM CPAN2*CNH TO SUPAN2
SUM CPAN3*CNH TO SUPAN3
SUM CPAN4*CNH TO SUPAN4
SUM CPRD*CNH TO SUPRD
SUM CPRD2*CNH TO SUPRD2
SUM CPRD3*CNH TO SUPRD3
SUM CPRD4*CNH TO SUPRD4
SUM COTROS*CNH TO SUOTROS
SUM COTROS2*CNH TO SUOTROS2
SUM COTROS3*CNH TO SUOTROS3
SUM COTROS4*CNH TO SUOTROS4
*SUM CVAL*CNH*CMH TO SUVAL

MPRI=SUPRI/4399   && MEDIA ESTRA. EST 1
MPRI2=SUPRI2/4399 && MEDIA ESTRA. EST 2
MPRI3=SUPRI3/4399 && MEDIA ESTRA. EST 3
MPRI4=SUPRI4/4399
MPAN=SUPAN/4399
MPAN2=SUPAN2/4399
MPAN3=SUPAN3/4399
MPAN4=SUPAN4/4399

MPRD=SUPRD/4399
MPRD2=SUPRD2/4399
MPRD3=SUPRD3/4399
MPRD4=SUPRD4/4399
MOTROS=SUOTROS/4399
MOTROS2=SUOTROS2/4399
MOTROS3=SUOTROS3/4399
MOTROS4=SUOTROS4/4399

USE RESULMES
APPEND BLANK
REPLACE RPRI WITH MPRI,RPAN WITH MPAN
REPLACE RPRI2 WITH MPRI2,RPRI3 WITH MPRI3
REPLACE RPRI4 WITH MPRI4
REPLACE RPAN2 WITH MPAN2,RPAN3 WITH MPAN3
REPLACE RPAN4 WITH MPAN4
**
REPLACE RPRD WITH MPRD,RPRD2 WITH MPRD2
REPLACE RPRD4 WITH MPRD4
REPLACE RPRD3 WITH MPRD3,ROTROS WITH MOTROS
REPLACE ROTROS2 WITH MOTROS2,ROTROS3 WITH MOTROS3
REPLACE ROTROS4 WITH MOTROS4
SET DECIMALS TO 4
```

```
BVEZ=BVEZ+1
BANDERA="R"
? BANDERA
*WAIT " BANDERA 2"
ENDDO
**********RETURN
? TSTRING(SECONDS()-TIEM)
? CHR(7)
#INCLUDE "ALGOC.PRG"
#INCLUDE "ALGOS.PRG"
```

P 15 (PROGRAMA 15 : CONVIERTE CASILLAS A SECCIONES)

```
*SECCION.PRG
*CONVIERTE A CASILLAS LAS SECCIONES DE  GUANAJUATO
*DE UNA BASE DE DATOS  GUANA2.DBF QUE CONTIENE CASILLAS A UNA BASE LLAMA DA
*GUANA.DBF QUE CONTENDRA LAS SECCIONES
USE GUANA2
GO TOP
DO WHILE .NOT. EOF()
 ? RECNO()
  M=MPIO
  D=DLOCAL
  S=SECCION
  U=UBICACION
  CPRI=0
  CPAN=0
  CPRD=0
  COTROS=0
  CVAL=0
  CLN=0
  DO WHILE SECCION=S
    CPRI=CPRI+PRI
    CPAN=CPAN+PAN
    CPRD=CPRD+PRD
    COTROS=COTROS+OTROS
    CVAL=CVAL+VALIDOS
     CLN=CLN+LISTA_NOM
     REG=RECNO()
     SKIP
   ENDDO
 USE GUANA
 APPEND BLANK
 REPLACE SECCION WITH S,UBICACION WITH U
 REPLACE PRI WITH CPRI, PAN WITH CPAN
 REPLACE PRD WITH CPRD, OTROS WITH COTROS
 REPLACE VALIDOS WITH CVAL, MPIO WITH M
 REPLACE DLOCAL WITH D,  LISTA_NOM WITH CLN
 USE GUANA2
 GO REG+1
ENDDO
```