UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

U.A.C.P. y P. del C.C.H.

I.I.M.A.S.

"DESARROLLO DE UN CONJUNTO DE RUTINAS PARA MICROCOMPUTADORAS PARA LA RESOLUCION DE PROBLEMAS BASICOS DE MUESTREO"

T E S I N A

QUE PARA OBTENER EL DIPLOMA DE

ESPECIALISTA EN ESTADISTICA APLICADA

P R E S E N T A

KOJI AKAGI

NOVIEMBRE, 1987

両親へ：

唯一の親孝行として

美和子さんへ：感謝の気持をこめて

友達ホセルスイへ：どうもありがとう

AGRADECIMIENTOS

Deseo expresar mi sincero agradecimiento a la profesora Adriana Ducoing Watty por su apoyo, su enseñanza, su critica y su consejo para la realización de este trabajo.

Asimismo quiero agradecer:

A la Secretaria de Relaciones Exteriores por su financiamiento.

Al P. de M. en C. Raymundo Eugenio Peralta Herrera por su apoyo de uso de la microcomputadora.

Al Biol. Jesús Sánchez Robles por su apoyo en la realización de este trabajo.

A las siguientes personas por su atenta revisión de este trabajo:

Dr. Rubén Hernández Cid

Act. Patricia Romero Mares

M. en C. Guillermo Baz Tellez

M. en C. Jorge Olguín Uribe

C O N T E N I D O

1. INTRODUCCION

En este trabajo se desarrollan una serie de rutinas de cómputo para la solución de problemas básicos de muestreo. Está elaborado escencialmente para servir como apoyo didáctico para un curso básico de muestreo, aunque puede ser útil para la resolución de problemas prácticos que no involucren diseños de muestra complejos.

Los programas se elaboraron en TURBO PASCAL (Versión 3.0) utilizando la microcomputadora PRINTAFORM con el sistema operativo MS-DOS (Versión 2.0) y para facilitar su uso son interactivos.

El paquete así como la documentación de los programas respectivos se encuentran a disposición de los usuarios en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas.

2. DESCRIPCION DE LOS PROCEDIMIENTOS DE ESTIMACION Y DETERMINACION DE TAMAÑO DE MUESTRA PARA LOS DIFERENTES ESQUEMAS DE MUESTREO

2.1. INTRODUCCION

En este capítulo, se describen los procedimientos para estimación de parámetros y determinación de tamaño de muestra para diferentes esquemas de muestreo.

Para cada parámetro a estimar se proporciona la siguiente información:

1) Estimador

2) Varianza del estimador o en su caso, la aproximación del error cuadrático medio del estimador

3) Estimador de la varianza del estimador o en su caso, el estimador de la aproximación del error cuadrático medio del estimador

4) Intervalos de confianza

   Se presentan procedimientos utilizando como aproximación a la distribución muestral de los estimadores la normal y la t de Student. Para los casos en que no es posible aproximar con algún modelo probabilístico a la distribución de un estimador, se utilizan los percentiles que se originan a partir de la desigualdad de Tchebycheff para la construcción de intervalos de confianza.

Para la determinación del tamaño de muestra, se proporcionan los siguientes procedimientos:

1) Determinación del tamaño de muestra que se requiera para tener una precisión y confianza prefijadas.

2) Determinación del tamaño de muestra que se requiera para obtener un coeficiente de variación del estimador prefijado.

En el caso del muestreo estratificado aleatorio, se incluyen además procedimientos óptimos de determinación de tamaño de muestra para cuando se utiliza asignación óptima o de Neyman. Se proporcionan dos alternativas:

1) Determinación del tamaño de muestra que minimice la varianza para un costo total fijo

2) Determinación del tamaño de muestra que minimice el costo total para una varianza fija

En el caso en el que el estimador para el que se quiere determinar el tamaño de muestra sea sesgado, se incluyen los procedimientos para asegurar que el sesgo sea despreciable.

Antes de mencionar los procedimientos de estimación y determinación de tamaño de muestra para cada esquema de muestreo por separado, se presenta una breve explicación acerca de los diferentes esquemas de muestreo que se incluyen en este trabajo y se discuten también brevemente los procedimientos de estimación que utilizan información auxiliar.

2.2. ESQUEMAS DE MUESTREO

Los métodos de muestreo que abarca este trabajo son:

1) Muestreo aleatorio simple sin reemplazo
2) Muestreo estratificado aleatorio
3) Muestreo con probabilidad proporcional al tamaño con reemplazo
4) Muestreo bietápico con selección en la primera etapa con probabilidad proporcional al tamaño con reemplazo y en la segunda etapa con muestreo aleatorio simple sin reemplazo

1) El muestreo aleatorio simple es el más usado cuando se tiene un marco de muestreo que especifique la manera de identificar cada unidad en la población.

Si una muestra de tamaño (n) es seleccionada de una población de tamaño (N), de tal forma que cada muestra posible del mismo tamaño tiene la misma probabilidad de ser seleccionada, a la muestra así obtenida se le conoce como muestra aleatoria simple. Cuando en el procedimiento de selección no se permite que un elemento aparezca en muestra más de una vez, a este procedimiento se le conoce como muestreo aleatorio simple sin reemplazo.

2) En el muestreo estratificado aleatorio, previamente a la selección, se divide a los elementos de la población en grupos exhaustivos y excluyentes llamados estratos. Estos grupos se forman de tal modo que se mejore la precisión de los estimadores (por ejemplo, los grupos deben ser homogéneos cuando se utilizan estimadores simples). En cada estrato, se obtiene por separado y de manera independiente una muestra aleatoria simple sin reemplazo de los elementos que lo conforman.

Existen varias razones por las que conviene utilizar este método de muestreo, pero las principales son las siguientes:

a) Reducir la variabilidad de los estimadores.

b) Disponibilidad diferencial de marcos de muestreo.

Cuando se tiene un buen marco para una parte de la población pero para la otra parte se tiene un marco impreciso, es conveniente utilizar este método de muestreo dividiendo los elementos en estratos según el marco de muestreo disponible y aplicar procedimientos de selección diferentes.

c) Costos de localización y levantamiento de información diferenciales.

El costo de obtener la información se puede reducir mediante la estratificación de los elementos de la población en grupos convenientes cuando el costo varía de un estrato a otro.

3) El muestreo con probabilidad proporcional al tamaño con reemplazo es aquel en el que se seleccionan los elementos de una población con probabilidad desigual pero proporcional al valor de una variable que está relacionada con la variable de interés (y) de manera proporcional directa.
Si se tiene proporcionalidad perfecta entre la variable (y) y la variable (x) (yi/xi = k, para toda i), la varianza resultante de los estimadores será cero.

4) Cuando no se tiene un marco de muestreo de elementos de la población o dichos elementos se encuentran muy dispersos, se forman (o se encuentran ya formados) grupos exhaustivos y excluyentes de elementos de preferencia heterogéneos, llamados conglomerados o unidades primarias de muestreo y se elabora un marco de muestreo para ellos. Cuando las unidades primarias son muchas, con un numero variable de elementos y homogéneos, es conveniente seleccionar algunas de ellas y en cada una de ellas seleccionan algunos elementos de la poblacion. A este procedimiento se le conoce como muestreo bietápico. Se pueden seleccionar unidades primarias y elementos con cualquier esquema de selección. En este trabajo se presenta el caso en el que las unidades primarias se seleccionan con probabilidad proporcional al total de unidades secundarias con reemplazo y los elementos con muestreo aleatorio simple sin reemplazo.

2.3. ESTIMACION DE RAZON Y REGRESION

Algunas veces se dispone de información acerca de una variable auxiliar (x) que se encuentra relacionada con la variable de interés (y). Esta información auxiliar se puede utilizar para construir estimadores llamados de razón y regresión, que bajo ciertas circunstancias, que a continuación se mencionan, tiene mayor precisión que los estimadores simples, a pesar de que son estimadores sesgados.

Los estimadores de razón se utilizan cuando la relación entre (x) y (y) es lineal a través del origen. Estos estimadores son más precisos que los estimadores simples cuando la correlación entre (x) y (y), ρ_{xy} con

$$\rho_{xy} > \frac{1}{2} \frac{CV(\bar{X})}{CV(\hat{\bar{Y}})}$$

y cuando existe proporcionalidad, es decir que la recta pasa por el origen. (CV es el coeficiente de variación)

Los estimadores de regresión se utilizan cuando la relación entre (x) y (y) es lineal aunque no pase por el origen. En muestras grandes estos estimadores son los más precisos para estimar una media o un total. (RaJ Des (1968) "Sampling Theory" McGRAW HILL Página 101)

Estos dos métodos de estimación pueden proporcionar estimadores más precisos que los simples en algunos casos que se mencionaron. Sin embargo debe tenerse en consideración el sesgo del estimador y el costo de obtención de la información sobre la variable (x).

Los estimadores de razón y regresión son sesgados pero consistentes, es decir la $E[\hat{\theta}] \neq \theta$ pero si $n \to \infty$, $E[\hat{\theta}] \to \theta$ y $V[\hat{\theta}] \to 0$.

Para medir la variabilidad alrededor de la media de un estimador sesgado se utiliza el error cuadrático medio:

$$ECM[\hat{\theta}] = V[\hat{\theta}] + (E[\hat{\theta}] - \theta)^2$$

$ECM[\hat{\theta}]$: Error cuadrático medio de $\hat{\theta}$
$V[\hat{\theta}]$: Varianza de $\hat{\theta}$
$E[\hat{\theta}]$: Valor esperado de $\hat{\theta}$
$E[\hat{\theta}] - \theta$: Sesgo de $\hat{\theta}$

Para construir intervalos de confianza es importante controlar la magnitud del sesgo. Posteriormente se presenta para cada esquema de selección algunas condiciones bajo las cuales se puede considerar el sesgo despreciable.

Finalmente, se debe considerar el costo de obtención de información sobre la variable auxiliar (x). Si la situación física sugiere el uso de la estimación de razón o regresión, el experimentador debe decidir si el incremento en la precisión del estimador de razón y regresión justifica el costo adicional.

2.4. MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

2.4.1. NOTACION

En esta sección se describe la notación necesaria para la descripción de los procedimientos de estimación y determinación del tamaño de muestra en el muestreo aleatorio simple sin reemplazo.

Valores poblacionales:

N : Número de elementos en la población

Y_i: Valor de la característica de interés para el i-ésimo elemento de la población

$\bar{Y}$: Media poblacional de la característica de interés

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

S_y^2 : Varianza poblacional de la característica de interés

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

Y : Total poblacional de la característica de interés

$$Y = \sum_{i=1}^{N} Y_i$$

X_i: Valor de la variable auxiliar para el i-ésimo elemento de la población

$\bar{X}$: Media poblacional de la variable auxiliar

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

S_x^2 : Varianza poblacional de la variable auxiliar

$$S_x^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}$$

X : Total poblacional de la variable auxiliar

$$X = \sum_{i=1}^{N} X_i$$

P : Proporción poblacional

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

R : Razón poblacional

$$R = \frac{Y}{X}$$

B : Coeficiente de regresión

$$B = \frac{\sum_{i=1}^{N} (Yi - \bar{Y})(Xi - \bar{X})}{\sum_{i=1}^{N} (Xi - \bar{X})^2}$$

ρxy : Coeficiente de correlación entre (x) y (y)

$$\rho xy = \frac{\sum_{i=1}^{N} (Xi - \bar{X})(Yi - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (Xi - \bar{X})^2 \sum_{i=1}^{N} (Yi - \bar{Y})^2}}$$

Valores muestrales

n : Tamaño de muestra

yi: Valor de la característica de interés para el i-ésimo elemento de la muestra

$\bar{y}$: Media muestral de la característica de interés

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

sy^2 : Varianza muestral de la característica de interés

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

xi: Valor de la variable auxiliar para el i-ésimo elemento de la muestra

$\bar{x}$: Media muestral de la variable auxiliar

$$\bar{x} = \frac{\sum_{i=1}^{n} xi}{n}$$

sx^2 : Varianza muestral de la variable auxiliar

$$sx^2 = \frac{\sum_{i=1}^{n} (xi - \bar{x})^2}{n - 1}$$

Para los estimadores, se usa el signo ^ encima de cada parámetro.

Otros:

$V[\hat{\theta}]$: Varianza de $\hat{\theta}$

$ECM[\hat{\theta}]$: Error cuadrático medio de $\hat{\theta}$

$P[\ A < \theta < B\ ]$: Probabilidad de que el parámetro θ caiga entre A y B.

2.4.2. PROCEDIMIENTOS DE ESTIMACION

En esta parte, se presentan los procedimientos de estimación de los siguientes parámetros:

1) Media
2) Total
3) Proporción
4) Razón

Se presentan estimadores simples para los cuatro parámetros y estimadores de razón y regresión para una media y un total.

1) Media

1.1) Estimador Simple

Estimador

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

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V[\hat{\bar{Y}}] = \left(1 - \frac{n}{N}\right)\frac{S_y^2}{n} \qquad S_y^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{\bar{Y}}] = \left(1 - \frac{n}{N}\right)\frac{s_y^2}{n} \qquad s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\bar{Y}})^2}{n-1}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

a) Si el tamaño de muestra es grande ($n \geq 30$ tamaño mínimo que empíricamente se ha considerado adecuado para la aproximación la distribución muestral con la distribución normal), se utiliza la aproximación normal.

$$P[\hat{\bar{Y}} - Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1-\alpha$$

Nota: $Z\ \alpha/2$ es el valor de Z obtenido en las tablas de la normal estandar que deja un area de $\alpha/2$.

b) Si el tamaño de muestra es pequeño ($n<30$), pero la variable (y) se distribuye normalmente, se usa la aproximación a la t de Student.

$$P[\hat{\bar{Y}} - t^{n-1}_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + t^{n-1}_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1-\alpha$$

Nota: $t^{n-1}\ \alpha/2$ es el valor de t obtenido en las tablas de la t de Student con (n-1) grados de libertad y que deja un area de $\alpha/2$.

c) Si el tamaño de muestra es pequeño ($n<30$) y la variable (y) no se distribuye normalmente, se utilizan los percentiles que se obtienen a partir de la desigualdad de Tchebycheff, es decir:

$$P[\left|\hat{\theta} - E[\hat{\theta}]\right| \leq K\sqrt{V[\hat{\theta}]}] = 1 - 1/K^2$$

y entonces el intervalo de confianza se construye de la siguiente manera:

$$P[\hat{\bar{Y}} - K\sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + K\sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1-\alpha = 1 - 1/K^2$$

1.2) Estimador de Razón

Estimador

$$\hat{\bar{Y}}r = \hat{R}\,\bar{X} \qquad\qquad \hat{R} = \frac{\sum_{i=1}^{n} yi}{\sum_{i=1}^{n} xi}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

El sesgo de $\hat{\bar{Y}}r$ está determinado por el sesgo de $\hat{R}$. Se puede demostrar que el sesgo estandarizado está acotado por el coeficiente de variación del estimador de la media de la variable auxiliar, es decir:

$$\frac{\left| E[\hat{R}] - R \right|}{\sqrt{V[\hat{R}]}} \leq CV(\hat{\bar{X}})$$

Se considera el sesgo despreciable si el $CV(\hat{\bar{X}}) \leq 0.1$.(Raj Des(1968) "Sampling Theory" McGRAW HILL N.Y Página 87)

Coeficiente de variación de $\hat{\bar{X}}$

$$\hat{CV}(\hat{\bar{X}}) = \frac{\sqrt{\hat{V}[\hat{\bar{X}}]}}{\hat{E}[\hat{\bar{X}}]} \qquad \hat{V}[\hat{\bar{X}}] = \left(1 - \frac{n}{N}\right) \frac{sx^2}{n}$$

Aproximación del error cuadrático medio del estimador

$$\dot{ECM}[\hat{\bar{Y}}r] = \left(1 - \frac{n}{N}\right) \frac{Su^2}{n} \qquad Su^2 = \frac{\sum_{i=1}^{N} (Yi - R\,Xi)^2}{N - 1}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\hat{\dot{ECM}}[\hat{\bar{Y}}r] = \left(1 - \frac{n}{N}\right) \frac{su^2}{n} \qquad su^2 = \frac{\sum_{i=1}^{n} (yi - \hat{R}\,xi)^2}{n - 1}$$

Intervalos de confianza al $(1-\alpha)\times100\%$

I) Si el $\hat{CV}(\hat{\bar{X}}) \leq 0.1$, se considera el sesgo despreciable. Por lo tanto el error cuadrático medio se considera conformado esencialmente por la varianza del estimador.

a) Si el tamaño de muestra es grande $(n>30)$, se utiliza la aproximación normal.

$$P\left[\hat{\bar{Y}}r - Z_{\alpha/2}\sqrt{\hat{ECM}[\hat{\bar{Y}}r]} < \bar{Y} < \hat{\bar{Y}}r + Z_{\alpha/2}\sqrt{\hat{ECM}[\hat{\bar{Y}}r]} \right] = 1 - \alpha$$

b) Si el tamaño de muestra es pequeño ($n<30$), se usa la desigualdad de Tchebycheff.

$$P[\,\hat{\bar{Y}}r - K\sqrt{ECM[\hat{\bar{Y}}r]} < \bar{Y} < \hat{\bar{Y}}r + K\sqrt{ECM[\hat{\bar{Y}}r]}\,] = 1 - \alpha \;\; = 1 - 1/K^2$$

II) Si el $\hat{CV}(\hat{\bar{X}}) > 0.1$, no se considera el sesgo despreciable. Por lo tanto no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

1.3) Estimador de Regresión

Estimador

$$\hat{\bar{Y}}lr = \bar{y} - \hat{B}\,(\bar{x} - \bar{X})$$

$$\hat{B} = \frac{\sum_{i=1}^{n} (yi - \bar{y})(xi - \bar{x})}{\sum_{i=1}^{n} (xi - \bar{x})^2}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} xi}{n} \qquad \bar{y} = \frac{\sum_{i=1}^{n} yi}{n}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

El sesgo de un estimador de regresión depende del sesgo del estimador del coeficiente de regresión. Para tamaños de muestra mayores a 50, este sesgo puede considerarse despreciable.(Cochran 1974 "Técnicas de Muestreo" Página 254)

I) Si el tamaño de muestra es grande ($n \geq 50$), se considera el sesgo despreciable y entonces se trabaja con la varianza del estimador que a continuación se describe. Nótese que ésta se debe utilizar únicamente con muestras grandes.

Varianza del estimador

$$V[\hat{\bar{Y}}lr] = \left(1 - \frac{n}{N}\right)(1 - \rho^2_{xy})\frac{Sy^2}{n}$$

$$\rho_{xy} = \frac{\sum_{i=1}^{N} (Xi - \bar{X})(Yi - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (Xi - \bar{X})^2 \sum_{i=1}^{N} (Yi - \bar{Y})^2}}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{\bar{Y}}lr] = \left(1 - \frac{n}{N}\right)\frac{1}{n(n-2)}\sum_{i=1}^{n}\left[(yi - \hat{\bar{Y}}) - \hat{B}\,(xi - \hat{\bar{X}})\right]^2$$

Intervalos de confianza al $(1-\alpha)\times100\%$

Como se supone que se trabaja con muestras grandes por la eliminación del sesgo, se utiliza la aproximación a la normal.

$$P\left[\hat{\bar{Y}}lr - Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}lr]} < \bar{Y} < \hat{\bar{Y}}lr + Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}lr]}\right] = 1 - \alpha$$

II) Si el tamaño de muestra es pequeño (n<50), no se considera el sesgo despreciable. Por lo tanto no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

2) Total

2.1) Estimador Simple

Estimador

$$\hat{Y} = N\,\frac{\sum_{i=1}^{n} yi}{n}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V[\hat{Y}] = N^2\left(1 - \frac{n}{N}\right)\frac{Sy^2}{n}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{Y}] = N^2\left(1 - \frac{n}{N}\right)\frac{sy^2}{n}$$

Intervalos de confianza al $(1-\alpha)\times100\%$

a) Si el tamaño de muestra es grande ($n \geq 30$), se utiliza la aproximación normal.

$$P[\hat{Y} - Z_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + Z_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]}] = 1 - \alpha$$

b) Si el tamaño de muestra es pequeño ($n<30$), pero la variable (y) se distribuye normalmente, se usa la aproximación a la t de Student.

$$P[\hat{Y} - t^{n-1}_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + t^{n-1}_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]}] = 1 - \alpha$$

c) Si el tamaño de muestra es pequeño ($n<30$) y la variable (y) no se distribuye normalmente, se usa la desigualdad de Tchebycheff.

$$P[\hat{Y} - K\sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + K\sqrt{\hat{V}[\hat{Y}]}] = 1 - \alpha = 1 - 1/K^2$$

2.2) Estimador de Razón

Estimador

$$\hat{Y}r = \hat{R}\, X \qquad \hat{R} = \frac{\sum_{i=1}^{n} yi}{\sum_{i=1}^{n} xi}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador de la media vía razón.

Aproximación del error cuadrático medio del estimador

$$ECM[\hat{Y}r] \doteq N^2 \left(1 - \frac{n}{N}\right) \frac{Su^2}{n}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\hat{ECM}[\hat{Y}r] \doteq N^2 \left(1 - \frac{n}{N}\right) \frac{su^2}{n}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

I) Si el $\hat{CV}(\hat{\bar{X}}) \leq 0.1$, se considera el sesgo despreciable.

a) Si el tamaño de muestra es grande ($n \geq 30$), se aproxima a la normal.

$$P[\,\hat{Yr} - Z_{\alpha/2}\sqrt{\hat{ECM}[\hat{Yr}]} < Y < \hat{Yr} + Z_{\alpha/2}\sqrt{\hat{ECM}[\hat{Yr}]}\,] = 1 - \alpha$$

b) Si el tamaño de muestra es pequeño ($n<30$), se usa la desigualdad de Tchebycheff.

$$P[\,\hat{Yr} - K\sqrt{\hat{ECM}[\hat{Yr}]} < Y < \hat{Yr} + K\sqrt{\hat{ECM}[\hat{Yr}]}\,] = 1 - \alpha = 1 - 1/K^2$$

II) Si el $\hat{CV}(\hat{\bar{X}}) > 0.1$, no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

2.3) Estimador de Regresión

Estimador

$$\hat{Ylr} = N\,(\bar{y} - \hat{B}(\bar{x} - \bar{X}))$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador de la media vía regresión.

I) Si el tamaño de muestra es grande ($n \geq 50$), se considera el sesgo despreciable. Se utiliza entonces la varianza del estimador para muestras grandes.

Varianza del estimador

$$\hat{V}[Ylr] = N^2\left(1 - \frac{n}{N}\right)\left(1 - \rho_{xy}^2\right)\frac{Sy^2}{n}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{Y}lr] = N^2 \left(1 - \frac{n}{N}\right) \frac{1}{n(n-2)} \sum_{i=1}^{n} [(yi - \bar{\hat{Y}}) - \hat{B}(xi - \bar{\hat{X}})]^2$$

Intervalos de confianza al $(1-\alpha) \times 100\%$

Se aproxima a la normal.

$$P[\hat{Y}lr - Z_{\alpha/2} \sqrt{\hat{V}[\hat{Y}lr]} < Y < \hat{Y}lr + Z_{\alpha/2} \sqrt{\hat{V}[\hat{Y}lr]}] = 1 - \alpha$$

II) Si el tamaño de muestra es pequeño ($n<50$), no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

3) Proporción

Estimador

$$\hat{P} = \frac{\sum_{i=1}^{n} yi}{n}$$

Nota: Este es un estimador insesgado.
yi toma el valor 1 cuando el i-ésimo elemento de la muestra posee la característica de interés y el valor 0 cuando el i-ésimo elemento de la muestra no posee la característica de interés.

Varianza del estimador

$$V[\hat{P}] = \left(1 - \frac{n}{N}\right) \frac{N}{N-1} \frac{PQ}{n} \qquad Q = 1 - P$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{P}] = \left(1 - \frac{n}{N}\right) \frac{\hat{P}\hat{Q}}{n-1} \qquad \hat{Q} = 1 - \hat{P}$$

Intervalos de confianza al $(1-\alpha)\times100\%$

Valores más pequeños de np para uso
de la aproximación normal
(Cochran 1974 "Técnicas de Muestreo" Página 88)

p	np	n
~0	80	∞
0.05	70	1400
0.1	60	600
0.2	40	200
0.3	24	80
0.4	20	50
0.5	15	30

a) Si el tamaño de muestra es mayor o igual a la n del cuadro que corresponde a la proporción estimada, entonces se puede utilizar la aproximación normal.

$$P[\hat{P} - Z_{\alpha/2}\sqrt{\hat{V}[\hat{P}]} < P < \hat{P} + Z_{\alpha/2}\sqrt{\hat{V}[\hat{P}]}] = 1 - \alpha$$

b) Si el tamaño de muestra es menor que la n del cuadro que corresponde a la proporción estimada, se utiliza la desigualdad de Tchebycheff.

$$P[\hat{P} - K\sqrt{\hat{V}[\hat{P}]} < P < \hat{P} + K\sqrt{\hat{V}[\hat{P}]}] = 1 - \alpha = 1 - 1/K^2$$

4) Razón

Estimador

$$\hat{R} = \frac{\sum_{i=1}^{n} yi}{\sum_{i=1}^{n} xi}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador de la media vía razón.

Aproximación del error cuadrático medio del estimador

$$\hat{ECM}[R] \doteq \frac{1}{\bar{X}^2}\left(1-\frac{n}{N}\right)\frac{Su^2}{n}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\hat{\hat{ECM}}[R] \doteq \frac{1}{\hat{\bar{X}}^2}\left(1-\frac{n}{N}\right)\frac{su^2}{n}$$

Intervalos de confianza al $(1-\alpha)\times100\%$

I) Si el $\hat{CV}(\hat{\bar{X}}) \leq 0.1$, se considera el sesgo despreciable.

a) Si el tamaño de muestra es grande $(n \geq 30)$, se aproxima a la normal.

$$P[\hat{R} - Z_{\alpha/2}\sqrt{\hat{\hat{ECM}}[R]} < R < \hat{R} + Z_{\alpha/2}\sqrt{\hat{\hat{ECM}}[R]}] \doteq 1-\alpha$$

b) Si el tamaño de muestra es pequeño $(n<30)$, se usa la desigualdad de Tchebycheff.

$$P[\hat{R} - K\sqrt{\hat{\hat{ECM}}[R]} < R < \hat{R} + K\sqrt{\hat{\hat{ECM}}[R]}] \doteq 1-\alpha \doteq 1 - 1/K^2$$

II) Si el $\hat{CV}(\hat{\bar{X}}) > 0.1$, no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

2.4.3. PROCEDIMIENTOS DE DETERMINACION DEL TAMAÑO DE MUESTRA

Se presentan los procedimientos para determinar el tamaño de muestra para cada caso anteriormente mencionado y en cada uno de ellos, se describen dos métodos alternativos para determinación del tamaño de muestra:

a) Determinación del tamaño de muestra que se requiere para tener una precisión y confianza prefijadas.

b) Determinación del tamaño de muestra que se requiere para obtener un coeficiente de variación del estimador prefijado.

1) Determinación del tamano de muestra para estimar una media poblacional

1.1) Mediante un estimador simple

a) Si se quiere fijar la precisión (BO) y la confianza ($1-\alpha$)

a.1) Si se puede suponer distribución normal para la variable (y)

i)
$$n' = \frac{Z^2_{\alpha/2}\, S_y^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: S_y^2 se estima con una prueba piloto o bien se aproxima con el conocimiento previo sobre la población.

ii) Si resultó n menor de 30. Sea $n = n_{i-1}$.

$$n_i' = \frac{\left(t^{\,n_{i-1}-1}_{\alpha/2}\right)^2 S_y^2}{BO^2} \qquad n_i = \frac{1}{\frac{1}{n_i'} + \frac{1}{N}}$$

Nota: El procedimiento ii) se repite hasta que $n_i = n_{i+1}$.

a.2) Si no se puede suponer distribución normal para la variable (y)

i)
$$n' = \frac{K^2\, S_y^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

donde K es el percentil resultante de la desigualdad de Tchebycheff.

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{S_y^2}{CVO\ \bar{Y}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: $\bar{Y}$ se estima con una prueba piloto o bien se aproxima con el conocimiento previo sobre la población.

1.2) Mediante un estimador de razón

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

i)
$$n' = \frac{Z^2_{\alpha/2}\, Su^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Su^2 se estima con una prueba piloto o bien se aproxima con el conocimiento previo sobre la población.

ii) Para asegurar que el sesgo sea despreciable, se calcula el tamaño de muestra que cumpla con que el $CV(\hat{\bar{X}}) = 0.10$.

$$n' = \frac{Sx^2}{(0.10)^2\, \bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Sx^2 y $\bar{X}$ se estiman con una prueba piloto o bien se aproximan con el conocimiento previo sobre la población.

Nota: Se escoge el mayor entre i) y ii).

b) Si se quiere fijar el coeficiente de variación (CVO)

i)
$$n' = \frac{Su^2}{CVO^2\, \bar{Y}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

ii) Para asegurar que el sesgo sea despreciable, se calcula el tamaño de muestra que cumpla con que el $CV(\hat{\bar{X}}) = 0.10$.

$$n' = \frac{Sx^2}{(0.10)^2\, \bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Se escoge el mayor entre i) y ii).

1.3) Mediante un estimador de regresión

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

$$n' = \frac{Z^2_{\alpha/2}(1-\rho^2_{xy})Sy^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: ρ_{xy} y Sy^2 se estiman con una prueba piloto o se aproximan con el conocimiento previo sobre la población.

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{(1-\rho^2_{xy})Sy^2}{CVO^2\,\bar{Y}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Si n resulta menor de 50, se tendra que aumentar a 50 para poder despreciar el sesgo.

2) Determinación del tamaño de muestra para estimar un total poblacional

2.1) Mediante un estimador simple

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

a.1) Si se puede suponer distribución normal para la variable (y)

i)
$$n' = \frac{Z^2_{\alpha/2}\,N^2\,Sy^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

ii) Si resultó n menor de 30. Sea $n = n_{i-1}$.

$$n_i' = \frac{(t^{n_{i-1}-1}_{\alpha/2})^2\,N^2\,Sy^2}{BO^2} \qquad n_i = \frac{1}{\frac{1}{n_i'} + \frac{1}{N}}$$

Nota: El procedimiento ii) se repite hasta que $n_i = n_{i+1}$.

a.2) Si no se puede suponer distribución normal para la variable (y)

$$n' = \frac{K^2\,N^2\,Sy^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{S_y^2}{CVO^2\ \bar{Y}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

2.2) Mediante un estimador de razón

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

i)

$$n' = \frac{Z^2_{\alpha/2}\ N^2\ S_u^2}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

ii) Para asegurar que el sesgo sea despreciable, se calcula el tamaño de muestra que cumpla con que el $CV(\hat{\bar{X}}) = 0.10$.

$$n' = \frac{S_x^2}{(0.10)^2\ \bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Se escoge el mayor entre i) y ii)

b) Si se quiere fijar el coeficiente de variación (CVO)

i)

$$n' = \frac{S_u^2}{CVO^2\ \bar{Y}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

ii) Para asegurar que el sesgo sea despreciable, se calcula el tamaño de muestra que cumpla con que el $CV(\hat{\bar{X}}) = 0.10$.

$$n' = \frac{S_x^2}{(0.10)^2\ \bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Se escoge el mayor entre i) y ii)

2.3) Mediante un estimador de regresión

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

$$n' = \frac{Z^2_{\alpha/2}\ N^2\ (1-\rho^2_{xy})S_y^2}{BO^2} \qquad n' = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{(1-\rho_{xy}^2)S_y^2}{CVO^2\,\bar{Y}^2} \qquad n' = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Si n resulta menor de 50, se tendra que aumentar a 50 para poder despreciar el sesgo.

3) Determinación del tamaño de muestra para estimar una proporción poblacional

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

i) Si la proporción no es muy pequeña (aqui se toma a $p>0.1$), se considera que la aproximación normal a la distribución muestral de P es buena.

$$n' = \frac{Z_{\alpha/2}^2 \frac{N}{N-1} PQ}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: P se estima con una prueba piloto o bien se aproxima con el conocimiento previo sobre la población.

ii) Si la proporción es demasiado pequeña $(p \leq 0.1)$

$$n' = \frac{K^2 \frac{N}{N-1} PQ}{BO^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

b) Si se quiere fijar el coeficiente de variación (CVO), que sería el caso alternativo cuando $P \leq 0.1$.

$$n' = \frac{\frac{N}{N-1} PQ}{CVO^2\,P^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

4) Determinación del tamaño de muestra para estimar una razón poblacional

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

i) $$n' = \frac{Z_{\alpha/2}^2\, Su^2}{BO^2\,\bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

ii) Para asegurar que el sesgo sea despreciable, se calcula el tamaño de muestra que cumpla con que el $CV(\hat{\bar{X}})= 0.10$.

$$n' = \frac{Sx^2}{(0.10)^2 \bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Se escoge el mayor entre i) y ii).

b) Si se quiere fijar el coeficiente de variación (CVO)

i)
$$n' = \frac{Su^2}{CVO^2 \bar{X}^2 R^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: R se estima con una prueba piloto o bien se aproxima con el conocimiento previo sobre la población.

ii) Para asegurar que el sesgo sea despreciable, se calcula el tamaño de muestra que cumpla con que el $CV(\hat{\bar{X}}) = 0.10$.

$$n' = \frac{Sx^2}{(0.10)^2 \bar{X}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

Nota: Se escoge el mayor entre i) y ii).

2.5. MUESTREO ESTRATIFICADO ALEATORIO

2.5.1. NOTACION

La notación que se utiliza para este esquema de muestreo es la siguiente:

Valores poblacionales

L : Número de estratos

Nh : Total de unidades en el h-ésimo estrato

N : Total de unidades en la población

$$N = \sum_{h=1}^{L} Nh$$

Wh : Proporción de la población en el h-ésimo estrato

$$Wh = \frac{Nh}{N}$$

Yhi : Valor de la característica de interés para el i-ésimo elemento del h-ésimo estrato en la población

$\overline{Yh}$: Media poblacional de la característica de interés en el h-ésimo estrato

$$\overline{Yh} = \frac{\sum_{i=1}^{Nh} Yhi}{Nh}$$

Yh : Total poblacional de la característica de interés en el h-ésimo estrato

$$Yh = \sum_{i=1}^{Nh} Yhi$$

Syh^2 : Varianza poblacional de la característica de interés en el h-ésimo estrato

$$Syh^2 = \frac{\sum_{i=1}^{Nh} (Yhi - \overline{Yh})^2}{Nh - 1}$$

$\overline{Y}$: Media poblacional de la característica de interés

$$\overline{Y} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{Nh} Yhi}{N}$$

Y : Total poblacional de la característica de interés.

$$Y = \sum_{h=1}^{L} \sum_{i=1}^{Nh} Yhi$$

Xhi : Valor de la variable auxiliar para el i-ésimo elemento del h-ésimo estrato en la población

$\bar{X}h$: Media poblacional de la variable auxiliar en el h-ésimo estrato

$$\bar{X}h = \frac{\sum_{i=1}^{Nh} Xhi}{Nh}$$

Xh : Total poblacional de la variable auxiliar en el h-ésimo estrato

$$Xh = \sum_{i=1}^{Nh} Xhi$$

Sxh^2 : Varianza poblacional de la variable auxiliar en el h-ésimo estrato

$$Sxh^2 = \frac{\sum_{i=1}^{Nh} (Xhi - \bar{X}h)^2}{Nh - 1}$$

$\bar{X}$: Media poblacional de la variable auxiliar

$$\bar{X} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{Nh} Xhi}{N}$$

X : Total poblacional de la variable auxiliar

$$X = \sum_{h=1}^{L} \sum_{i=1}^{Nh} Xhi$$

Ph : Proporción poblacional en el h-ésimo estrato

$$Ph = \frac{Yh}{Nh}$$

Rh : Razón poblacional en el h-ésimo estrato

$$Rh = \frac{Yh}{Xh}$$

Bh : Coeficiente de regresión poblacional en el h-ésimo estrato

$$Bh = \frac{\sum_{i=1}^{Nh} (Yhi - \bar{Y}h)(Xhi - \bar{X}h)}{\sum_{i=1}^{Nh} (Xhi - \bar{X}h)^2}$$

ρ_{xyh} : Coeficiente de correlación entre (x) y (y)

$$\rho_{xyh} = \frac{\sum_{i=1}^{Nh} (Yhi - \bar{Y}h)(Xhi - \bar{X}h)}{\sqrt{\sum_{i=1}^{Nh} (Xhi - \bar{X}h)^2 \sum_{i=1}^{Nh} (Yhi - \bar{Y}h)^2}}$$

Valores muestrales

nh : Tamaño de muestra en el h-ésimo estrato

n : Tamaño de muestra total

$$n = \sum_{h=1}^{L} nh$$

yhi : Valor de la característica de interés para el i-ésimo elemento de la muestra del h-ésimo estrato

$\bar{y}h$: Media muestral de la característica de interés en el h-ésimo estrato

$$\bar{y}h = \frac{\sum_{i=1}^{nh} yhi}{nh}$$

syh^2 : Varianza muestral de la característica de interés en el h-ésimo estrato

$$syh^2 = \frac{\sum_{i=1}^{nh} (yhi - \bar{y}h)^2}{nh - 1}$$

xhi : Valor de la variable auxiliar para el i-ésimo elemento de la muestra del h-ésimo estrato

$\bar{x}h$: Media muestral de la variable auxiliar en el h-ésimo estrato

$$\bar{x}h = \frac{\sum_{i=1}^{nh} xhi}{nh}$$

s_{xh}^2 : Varianza muestral de la variable auxiliar en el h-ésimo estrato

$$s_{xh}^2 = \frac{\sum_{i=1}^{nh} (x_{hi} - \bar{x}_h)^2}{nh - 1}$$

Se utiliza el signo ^ para estimadores. La notación para la varianza, el error cuadrático medio y la probabilidad es la misma que para el muestreo aleatorio simple.

2.5.2. PROCEDIMIENTOS DE ESTIMACION

En esta parte, se presentan los procedimientos de estimación de los siguientes parámetros:

1) Media
2) Total
3) Proporción
4) Razón

Se presentan estimadores simples para los cuatro parámetros y estimadores de razón y regresión para una media y un total.

Para este tipo de diseño se presentan además dos opciones para la obtención de estimadores de razón y regresión: estimadores combinados y estimadores separados.

1) Media

1.1) Estimador Simple

Estimador

$$\hat{\bar{Y}} = \sum_{h=1}^{L} W_h \hat{\bar{Y}}_h \qquad W_h = \frac{Nh}{N} \qquad \hat{\bar{Y}}_h = \frac{\sum_{i=1}^{nh} y_{hi}}{nh}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V(\hat{\bar{Y}}) = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{S_{yh}^2}{nh} \qquad S_{yh}^2 = \frac{\sum_{i=1}^{Nh} (Y_{hi} - \bar{Y}_h)^2}{Nh - 1}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{\bar{Y}}] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{yh}^2}{n_h} \qquad s_{yh}^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \hat{\bar{Y}}_h)^2}{n_h - 1}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

a) Si el tamaño de muestra por estrato es grande en todos los estratos (regla empírica $n_h \geq 30$, para toda h), se utiliza la aproximación a la normal de la distribución muestral del estimador.

$$P[\hat{\bar{Y}} - Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1 - \alpha$$

b) Si Por lo menos un tamaño de muestra por estrato es pequeño ($n_h<30$, para alguna h) pero la variable (y) se distribuye normalmente en todos los estratos, se usa la aproximación a la t de Student con los grados de libertad corregidos.

$$P[\hat{\bar{Y}} - t_{\alpha/2}^{n_e}\sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + t_{\alpha/2}^{n_e}\sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1 - \alpha$$

Nota: Los grados de libertad de la t de Student se calculan de la manera siguiente:(Cochran 1974 "Técnicas de Muestreo" Página 136).

$$n_e = \frac{\left(\sum_{h=1}^{L} g_h s_h^2\right)^2}{\sum_{h=1}^{L} \frac{g_h^2 s_h^4}{n_h - 1}} \qquad g_h = \frac{N_h (N_h - n_h)}{n_h}$$

c) Si por lo menos un tamaño de muestra por estrato es pequeño ($n_h<30$, para alguna h) y la variable (y) no se distribuye normalmente por lo menos en un estrato, se usa la desigualdad de Tchebycheff.

$$P[\hat{\bar{Y}} - K\sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + K\sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1 - \alpha = 1 - 1/K^2$$

1.2) Estimador de Razón

1.2.1) Estimador Separado

Estimador

$$\hat{\bar{Y}}_{rs} = \sum_{h=1}^{L} W_h \hat{R}_h \bar{X}_h \qquad \hat{R}_h = \frac{\sum_{i=1}^{nh} y_{hi}}{\sum_{i=1}^{nh} x_{hi}}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Como el estimador separado es una media ponderada de los estimadores de razón por estrato, su sesgo depende de los sesgos de los estimadores de razón en cada estrato.

Suponiendo que el sesgo y la varianza del estimador de razón en cada estrato son los mismos, digamos

$$Sesgo(\hat{\bar{Y}}_{rh}) = Sesgo0 \text{ y } V[\hat{\bar{Y}}_{rh}] = V0,$$

se tiene entonces que:

$$\frac{Sesgo(\hat{\bar{Y}}_{rs})}{\sqrt{V[\hat{\bar{Y}}_{rs}]}} = \frac{\left|E[\hat{\bar{Y}}_{rs}] - \bar{Y}\right|}{\sqrt{V[\hat{\bar{Y}}_{rs}]}} = \frac{L \;\; Sesgo0}{\sqrt{L}\,\sqrt{V0}}$$

y como el sesgo estandarizado de un estimador de razón se encuentra acotado por el coeficiente de variación del estimador de la media de la variable auxiliar (Ver 2.4.2) se tiene entonces que

$$\frac{Sesgo\ (\hat{\bar{Y}}_{rs})}{\sqrt{V[\hat{\bar{Y}}_{rs}]}} \leqslant \frac{L}{\sqrt{L}}\, CV(\hat{\bar{X}}_{ho})$$

Por lo tanto el sesgo de este estimador no se puede despreciar si el número de estratos es muy grande a menos que los sesgos de los estimadores por estrato sean muy pequeños. Cochran(1974 "Técnicas de Muestreo" Página 222) propone que si $\sqrt{L}\ CV(\hat{\bar{X}}_{ho}) \leqslant 0.3$, se desprecie el sesgo de $\hat{\bar{Y}}_{rs}$.

Como en la práctica es difícil que el sesgo y la varianza de los estimadores de razón por estrato sean los mismos en todos los estratos, se propone calcular $\sqrt{L}\ CV(\hat{\bar{X}}_{ho})$ tomando $CV(\hat{\bar{X}}_{ho}) = \underset{h}{Max}\ \{ CV(\hat{\bar{X}}_h) \}$ y si resulta $\leqslant 0.3$ despreciar el sesgo de $\hat{\bar{Y}}_{rs}$. En caso contrario se sugiere aumentar el tamaño de muestra por estrato.

Aproximación del error cuadrático medio del estimador

$$ECM[\hat{\bar{Y}}_{rs}] \doteq \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{uh}^2}{n_h}$$

$$S_{uh}^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - R_h X_{hi})^2}{N_h - 1}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\widehat{ECM}[\hat{\bar{Y}}_{rs}] \doteq \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{uh}^2}{n_h}$$

$$s_{uh}^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \hat{R}_h x_{hi})^2}{n_h - 1}$$

Intervalos de confianza al $(1-\alpha) \times 100\%$

I) Si $\sqrt{L}\, \widehat{CV}(\hat{\bar{X}}_{ho}) \leq 0.3$, se considera el sesgo despreciable.

a) Si el tamaño de muestra por estrato es grande en todos los estratos ($n_h \geq 30$, para toda h), se aproxima a la normal.

$$P[\hat{\bar{Y}}_{rs} - Z_{\alpha/2} \sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rs}]} < \bar{Y} < \hat{\bar{Y}}_{rs} + Z_{\alpha/2} \sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rs}]}] = 1 - \alpha$$

b) Si por lo menos un tamaño de muestra por estrato es pequeño ($n_h < 30$, para alguna h), se usa la desigualdad de Thebycheff.

$$P[\hat{\bar{Y}}_{rs} - K \sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rs}]} < \bar{Y} < \hat{\bar{Y}}_{rs} + K \sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rs}]}] = 1 - \alpha = 1 - 1/K^2$$

II) Si $\sqrt{L}\, \widehat{CV}(\hat{\bar{X}}_{ho}) > 0.3$, no se considera el sesgo despreciable.
Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra para algunos estratos.

1.2.2) Estimador Combinado

Estimador

$$\hat{\bar{Y}}_{rc} = \hat{R}_c \bar{X} \qquad \hat{R}_c = \frac{\sum_{h=1}^{L} \hat{Y}_h}{\sum_{h=1}^{L} \hat{X}_h}$$

$$\hat{Y}_h = N_h \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

$$\hat{X}_h = N_h \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

El sesgo de $\hat{Y}_{rc}$ esta determinado por $\hat{R}_c$. Se puede demostrar que $\frac{\text{Sesgo } \hat{R}_c}{\sqrt{V[\hat{R}_c]}} \leqslant CV(\hat{\bar{X}}_{st})$.

$$CV(\hat{\bar{X}}_{st}) = \frac{\sqrt{V[\hat{\bar{X}}_{st}]}}{E[\hat{\bar{X}}]} \qquad V[\hat{\bar{X}}_{st}] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{xh}^2}{n_h}$$

Si $CV(\hat{\bar{X}}_{st}) \leqslant 0.1$, se considera el sesgo despreciable. (Cochran 1974 "Técnicas de Muestreo" Página 223)

Aproximación del error cuadrático medio del estimador

$$ECM[\hat{\bar{Y}}_{rc}] \doteq \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{uh}^2}{n_h}$$

$$S_{uh}^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - R\, X_{hi})^2}{N_h - 1}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\widehat{ECM}[\hat{\bar{Y}}_{rc}] \doteq \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{uh}^2}{n_h}$$

$$s_{uh}^2 = \frac{\sum_{i=1}^{n_h} [(y_{hi} - \bar{y}_h) - \hat{R}_c(x_{hi} - \bar{x}_h)]^2}{n_h - 1}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

I) Si $\widehat{CV}(\hat{\bar{X}}_{st}) \leq 0.1$, se considera el sesgo despreciable.

a) Si el tamaño de muestra total es grande $(n \geq 30)$, se aproxima a la normal.

$$P[\hat{\bar{Y}}_{rc} - Z_{\alpha/2}\sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rc}]} < \bar{Y} < \hat{\bar{Y}}_{rc} + Z_{\alpha/2}\sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rc}]}\,] = 1 - \alpha$$

b) Si el tamaño de muestra total es pequeño $(n<30)$, se usa la desigualdad de Tchebycheff.

$$P[\hat{\bar{Y}}_{rc} - K\sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rc}]} < \bar{Y} < \hat{\bar{Y}}_{rc} + K\sqrt{\widehat{ECM}[\hat{\bar{Y}}_{rc}]}\,] = 1 - \alpha = 1 - 1/K^2$$

II) Si $\widehat{CV}(\hat{\bar{X}}_{st}) > 0.1$, no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

1.2.3) Elección entre los estimadores separados y combinados

a) Si las razones por estrato, $\hat{R}_h$, varían considerablemente entre estratos, se deberá utilizar un estimador separado siempre que el número de estratos no sea muy grande y que los tamaños de muestra en cada estrato sean suficientemente grandes como para reducir el sesgo de $\hat{R}_h$ para toda h.

b) En caso de que no se cumpla alguna de las condiciones arriba mencionadas, se recomienda utilizar un estimador combinado.

1.3) Estimador de Regresión

1.3.1) Estimador Separado

Estimador

$$\hat{Y}_{lrs} = \sum_{h=1}^{L} W_h\left(\bar{y}_h - \hat{B}_h(\bar{x}_h - \bar{X}_h)\right)$$

$$\hat{B}_h = \frac{\sum_{i=1}^{n_h}(x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{\sum_{i=1}^{n_h}(x_{hi} - \bar{x}_h)^2} \qquad \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \qquad \bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Como el estimador separado es una media ponderada de los estimadores de regresión por estrato, su sesgo depende de los sesgos de los estimadores de regresión en cada estrato. Por lo tanto se considera el sesgo despreciable si el tamaño de muestra para cada estrato es mayor de 50, en cuyo caso la varianza es la siguiente:

I) Si el tamaño de muestra por estrato es grande en todos los estratos ($n_h \geq 50$, para toda h), se considera el sesgo despreciable y se usa la varianza del estimador para muestras grandes.

Varianza del estimador

$$V[\hat{Y}_{lrs}] = \sum_{i=1}^{L} W_h^2\left(1 - \frac{n_h}{N_h}\right)\left(1 - \rho_{xyh}^2\right)\frac{S_{yh}^2}{n_h}$$

$$\rho_{xyh} = \frac{\sum_{i=1}^{N_h}(X_{hi} - \bar{X}_h)(Y_{hi} - \bar{Y}_h)}{\sqrt{\sum_{i=1}^{N_h}(X_{hi} - \bar{X}_h)^2 \sum_{i=1}^{N_h}(Y_{hi} - \bar{Y}_h)^2}}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{\bar{Y}}lrs] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{1}{nh(nh-2)} \sum_{i=1}^{nh} [(yhi - \bar{y}h) - \hat{B}h\,(xhi - \bar{x}h)]^2$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

Se aproxima a la normal.

$$P[\ \hat{\bar{Y}}lrs - Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}lrs]} < \bar{Y} < \hat{\bar{Y}}lrs + Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}lrs]}\ ] = 1 - \alpha$$

II) Si por lo menos un tamaño de muestra por estrato es pequeño (nh<50,para alguna h), no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra para algunos estratos.

1.3.2) Estimador Combinado

Estimador

$$\hat{\bar{Y}}lrc = \hat{\bar{Y}}st - \hat{B}c\ (\ \hat{\bar{X}}st - \bar{X}\ )$$

$$\hat{B}c = \frac{\displaystyle\sum_{h=1}^{L} \frac{W_h^2 \left(1 - \frac{nh}{Nh}\right)}{nh(nh-1)}\, sxyh}{\displaystyle\sum_{h=1}^{L} \frac{W_h^2 \left(1 - \frac{nh}{Nh}\right)}{nh(nh-1)}\, sxxh}$$

$$sxyh = \sum_{i=1}^{nh} (xhi - \bar{x}h)(yhi - \bar{y}h)$$

$$sxxh = \sum_{i=1}^{nh} (xhi - \bar{x}h)^2$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Para determinar el sesgo de este estimador debe considerarse el tipo de asignación que se utilizó y la varianza residual en cada estrato. Cochran (1974 "Técnicas de Muestreo" Página 261) describe que si la varianza residual es constante entre estratos y la asignación proporcional, el sesgo del estimador es del orden de 1/n,
donde

$$n = \sum_{h=1}^{L} nh.$$

Por cuestiones prácticas en este trabajo, se considera el sesgo de este estimador despreciable si n es mayor de 50. Sin embargo no debe perderse de vista que dicho sesgo depende de una serie de consideraciones.(varianza residual, etc.)

I) Si el tamaño de muestra total es grande ($n \geq 50$), se considera el sesgo despreciable y se usa la varianza del estimador para muestras grandes.

Varianza del estimador

$$V[\hat{\bar{Y}}lrc] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{(Syh^2 + Bc^2\, Sxh^2 - 2Bc\, Cov(x,y)h)}{nh}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{\bar{Y}}lrc] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{1}{nh(nh-1)} \sum_{i=1}^{nh} [(yhi - \bar{y}h) - \hat{B}c\,(xhi - \bar{x}h)]^2$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

Se aproxima a la normal.

$$P[\hat{\bar{Y}}lrc - Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}lrc]} < \bar{Y} < \hat{\bar{Y}}lrc + Z_{\alpha/2}\sqrt{\hat{V}[\hat{\bar{Y}}lrc]}\,] = 1 - \alpha$$

II) Si el tamaño de muestra total es pequeño($n<50$), no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

1.3.3) Elección entre los estimadores separados y combinados

No se pueden dar criterios absolutos para decidir cual es mejor. Sin embargo si se tiene confianza en que las regresiones son lineales en todos los estratos entonces:

a) Si Bh parece ser el mismo entre estratos, debe preferirse el estimador combinado.

b) Si Bh's son diferentes entre estratos, se recomienda el estimador separado.

2) Total

2.1) Estimador Simple

Estimador

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h \qquad \hat{Y}_h = N_h \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V[\hat{Y}] = \sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{yh}^2}{n_h} \qquad S_{yh}^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2}{N_h - 1}$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{Y}] = \sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{yh}^2}{n_h} \qquad s_{yh}^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \hat{\bar{Y}}_h)^2}{n_h - 1}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

a) Si el tamaño de muestra por estrato es grande en todos los estratos ($n_h \geq 30$, para toda h), se aproxima a la normal.

$$P\left[\hat{Y} - Z_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + Z_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]}\right] = 1 - \alpha$$

b) Si por lo menos un tamaño de muestra por estrato es pequeño ($nh<30$, para alguna h) pero la variable (y) se distribuye normalmente en todos los estratos, se usa la aproximación a la t de Student.

$$P[\ \hat{Y} - t^{ne}_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + t^{ne}_{\alpha/2}\sqrt{\hat{V}[\hat{Y}]}\ ] = 1 - \alpha$$

Nota: Los grados de libertad de la t de Student se calculan de la manera siguiente:

$$ne = \frac{\left(\sum_{h=1}^{L} gh\ sh^2\right)^2}{\sum_{h=1}^{L} \frac{gh^2\ sh^4}{nh - 1}} \qquad gh = \frac{Nh\ (Nh - nh)}{nh}$$

c) Si por lo menos un tamaño de muestra por estrato es pequeño ($nh<30$, para alguna h) y la variable (y) no se distribuye normalmente por lo menos en un estrato, se usa la desigualdad de Tchebycheff.

$$P[\ \hat{Y} - K\sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + K\sqrt{\hat{V}[\hat{Y}]}\ ] = 1 - \alpha = 1 - 1/K^2$$

2.2) Estimador de Razón

2.2.1) Estimador Separado

Estimador

$$\hat{Y}rs = \sum_{h=1}^{L} \hat{R}h\ Xh \qquad \hat{R}h = \frac{\sum_{i=1}^{nh} yhi}{\sum_{i=1}^{nh} xhi}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador separado de la media vía razón.

Aproximación del error cuadrático medio del estimador

$$\mathrm{ECM}[\hat{Y}rs] \doteq \sum_{h=1}^{L} N_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{Suh^2}{nh}$$

$$Suh^2 = \frac{\sum_{i=1}^{Nh} (Yhi - Rh\, Xhi)^2}{Nh - 1}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\hat{\mathrm{E}}\mathrm{CM}[\hat{Y}rs] \doteq \sum_{h=1}^{L} N_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{suh^2}{nh}$$

$$suh^2 = \frac{\sum_{i=1}^{nh} (yhi - \hat{R}h\, xhi)^2}{nh - 1}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

I) Si $\sqrt{L}\ \hat{C}V(\hat{\bar{X}}ho) \leq 0.3$, se considera el sesgo despreciable.

a) Si el tamaño de muestra por estrato es grande en todos los estratos ($nh \geq 30$, para toda h), se aproxima a la normal.

$$P[\hat{Y}rs - Z_{\alpha/2}\sqrt{\hat{\mathrm{E}}\mathrm{CM}[\hat{Y}rs]} < Y < \hat{Y}rs + Z_{\alpha/2}\sqrt{\hat{\mathrm{E}}\mathrm{CM}[\hat{Y}rs]}] = 1 - \alpha$$

b) Si por lo menos un tamaño de muestra por estrato es pequeño ($nh < 30$, para alguna h), se usa la desigualdad de Tchebycheff

$$P[\hat{Y}rs - K\sqrt{\hat{\mathrm{E}}\mathrm{CM}[\hat{Y}rs]} < Y < \hat{Y}rs + K\sqrt{\hat{\mathrm{E}}\mathrm{CM}[\hat{Y}rs]}] = 1 - \alpha = 1 - 1/K^2$$

II) Si $\sqrt{L}\ \hat{C}V(\hat{\bar{X}}ho) > 0.3$, no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra para algunos estratos.

2.2.2) Estimador Combinado

Estimador

$$\hat{Y}_{rc} = \hat{R}_c X \qquad \hat{R}_c = \frac{\sum_{h=1}^{L} \hat{Y}_h}{\sum_{h=1}^{L} \hat{X}_h}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador combinado de la media vía razón.

Aproximación del error cuadrático medio del estimador

$$ECM[\hat{Y}_{rc}] \doteq \sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{uh}^2}{n_h}$$

$$S_{uh}^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - R\, X_{hi})^2}{N_h - 1}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\hat{ECM}[\hat{Y}_{rc}] \doteq \sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{uh}^2}{n_h}$$

$$s_{uh}^2 = \frac{\sum_{i=1}^{n_h} [(y_{hi} - \bar{y}_h) - \hat{R}_c(x_{hi} - \bar{x}_h)]^2}{n_h - 1}$$

Intervalos de confianza al $(1-\alpha)\times100\%$

I) Si $CV(\hat{\bar{X}}_{st}) \leq 0.1$, se considera el sesgo despreciable.

a) Si el tamaño de muestra total es grande ($n \geq 30$), se aproxima a la normal.

$$P[\hat{Y}rc - Z_{\alpha/2}\sqrt{\hat{E}\hat{C}M[Yrc]} < Y < \hat{Y}rc + Z_{\alpha/2}\sqrt{\hat{E}\hat{C}M[Yrc]}\ ] = 1-\alpha$$

b) Si el tamaño de muestra total es pequeño ($n<30$), se usa la desigualdad de Tchebycheff.

$$P[\hat{Y}rc - K\sqrt{\hat{E}\hat{C}M[Yrc]} < Y < \hat{Y}rc + K\sqrt{\hat{E}\hat{C}M[Yrc]}\ ] = 1 - \alpha = 1 - 1/K^2$$

II) Si $\hat{CV}(\hat{\bar{X}}st) > 0.1$, no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

2.2.3) Elección entre los estimadores separados y combinados

Las condiciones para elegir son las mismas que para el estimador de la media vía razón.

2.3) Estimador de Regresion

2.3.1) Estimador Separado

Estimador

$$\hat{Y}lrs = \sum_{h=1}^{L} N_h\ (\ \bar{y}h - \hat{B}h\ (\ \bar{x}h\ - \bar{X}h\ ))$$

$$\hat{B}h = \frac{\sum_{i=1}^{nh}(xhi - \bar{x}h)(yhi - \bar{y}h)}{\sum_{i=1}^{nh}(xhi - \bar{x}h)^2} \qquad \bar{y}h = \frac{\sum_{i=1}^{nh} yhi}{nh} \qquad \bar{x}h = \frac{\sum_{i=1}^{nh} xhi}{nh}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador separado de la media vía regresión.

I) Si el tamaño de muestra por estrato es grande en todos los estratos ($nh \geq 30$, para toda h), se considera el sesgo despreciable y entonces se usa la varianza del estimador para muestras grandes.

Varianza del estimador

$$\hat{V}[Ylrs] = \sum_{i=1}^{L} N_h^2 \left(1 - \frac{nh}{Nh}\right)(1 - \rho_{xyh}^2) \frac{Syh^2}{nh}$$

$$\rho_{xyh} = \frac{\sum_{i=1}^{Nh} (Xhi - \bar{X}h)(Yhi - \bar{Y}h)}{\sqrt{\sum_{i=1}^{Nh} (Xhi - \bar{X}h)^2 \sum_{i=1}^{Nh} (Yhi - \bar{Y}h)^2}}$$

Estimador de la varianza del estimador

$$\hat{\hat{V}}[Ylrs] = \sum_{h=1}^{L} N_h^2 \left(1 - \frac{nh}{Nh}\right)\frac{1}{nh(nh-2)} \sum_{i=1}^{nh} [(yhi - \bar{y}h) - \hat{B}h\,(xhi - \bar{x}h)]^2$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

Se aproxima a la normal.

$$P[\,\hat{Y}lrs - Z_{\alpha/2}\sqrt{\hat{\hat{V}}[Ylrs]} < Y < \hat{Y}lrs + Z_{\alpha/2}\sqrt{\hat{\hat{V}}[Ylrs]}\,] = 1 - \alpha$$

II) Si por lo menos un tamaño de muestra es pequeño (nh<50, para alguna h), no se considera el sesgo despreciable. Por lo tanto no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra para algunos estratos.

2.3.2) Estimador Combinado

Estimador

$$\hat{Y}lrc = N\,(\hat{\bar{Y}}st - \hat{B}c\,(\hat{\bar{X}}st - \bar{X}\,))$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador combinado de la media vía regresión.

I) Si el tamaño de muestra total es grande ($n \geq 50$), se considera el sesgo despreciable y se usa la varianza del estimador para muestras grandes.

Varianza del estimador

$$V[\hat{Y}lrc] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{Syh^2 + Bc^2 Sxh^2 - 2Bc\, Cov(x,y)h}{nh}$$

Estimador de varianza de estimador

$$\hat{V}[\hat{Y}lrc] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{1}{nh(nh-1)} \sum_{i=1}^{nh} [(yhi - \bar{y}h) - \hat{B}c\,(xhi - \bar{x}h)]^2$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

Se aproxima a la normal.

$$P[\hat{Y}lrc - Z_{\alpha/2}\sqrt{\hat{V}[\hat{Y}lrc]} < Y < \hat{Y}lrc + Z_{\alpha/2}\sqrt{\hat{V}[\hat{Y}lrc]}] = 1 - \alpha$$

II) Si el tamaño de muestra total es pequeño ($n<50$), no se considera el sesgo despreciable. Por lo tanto no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

3) Proporción

Estimador

$$\hat{P} = \sum_{h=1}^{L} W_h \hat{P}_h \qquad \hat{P}h = \frac{\sum_{i=1}^{nh} yhi}{nh}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V[\hat{P}] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{nh}{Nh}\right) \frac{Nh}{Nh - 1} \frac{Ph\ Qh}{nh} \qquad Qh = 1 - Ph$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{P}] = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{P}_h \hat{Q}_h}{n_h - 1}$$

Intervalos de confianza al $(1-\alpha) \times 100\%$

Valores más pequeños de np para uso de la aproximación normal

p	np	n
~0	80	∞
0.05	70	1400
0.1	60	600
0.2	40	200
0.3	24	80
0.4	20	50
0.5	15	30

a) Si el tamaño de muestra por estrato es mayor o igual a la n del cuadro que corresponde a la proporción estimada en ese estrato y ésto occure en todos los estratos, se utiliza la aproximación a la normal.

$$P[\hat{P} - Z_{\alpha/2} \sqrt{\hat{V}[\hat{P}]} < P < \hat{P} + Z_{\alpha/2} \sqrt{\hat{V}[\hat{P}]}] = 1 - \alpha$$

b) Si por lo menos un tamaño de muestra por estrato es menor que la n del cuadro que corresponde a la proporción estimada en ese estrato, se usa la desigualdad de Tchebycheff.

$$P[\hat{P} - K \sqrt{\hat{V}[\hat{P}]} < P < \hat{P} + K \sqrt{\hat{V}[\hat{P}]}] = 1 - \alpha = 1 - 1/K^2$$

4) Razón

Estimador

$$\hat{R} = \frac{\sum_{h=1}^{L} \hat{Y}_h}{\sum_{h=1}^{L} \hat{X}_h} \qquad \hat{Y}_h = N_h \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \qquad \hat{X}_h = N_h \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

Nota: Este es un estimador sesgado pero consistente.

Sesgo del estimador

Las consideraciones acerca del sesgo de este estimador y su control son las mismas que para el estimador combinado de la media vía razón.

Aproximación del error cuadrático medio del estimador

$$ECM[\hat{R}] = \frac{1}{X^{2}} \sum_{h=1}^{L} \left(1 - \frac{nh}{Nh}\right) \frac{Nh^{2}\, Suh^{2}}{nh}$$

$$Suh^{2} = \frac{\sum_{i=1}^{Nh} (Yhi - R\, Xhi)^{2}}{Nh - 1}$$

Estimador de la aproximación del error cuadrático medio del estimador

$$\hat{ECM}[\hat{R}] = \frac{1}{\hat{X}^{2}} \sum_{h=1}^{L} \left(1 - \frac{nh}{Nh}\right) \frac{Nh^{2}\, suh^{2}}{nh}$$

$$suh^{2} = \frac{\sum_{i=1}^{nh} [(yhi - \bar{y}h) - \hat{R}(xhi - \bar{x}h)]^{2}}{nh - 1}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

I) Si $\hat{CV}(\hat{\bar{X}}st) \leq 0.1$, se considera el sesgo despreciable.

a) Si el tamaño de muestra total es grande $(n \geq 30)$, se aproxima a la normal.

$$P[\hat{R} - Z_{\alpha/2} \sqrt{\hat{ECM}[\hat{R}]} < R < \hat{R} + Z_{\alpha/2} \sqrt{\hat{ECM}[\hat{R}]}\ ] = 1 - \alpha$$

b) Si el tamaño de muestra total es pequeño $(n<30)$, se usa la desigualdad de Tchebycheff.

$$P[\hat{R} - K \sqrt{\hat{ECM}[\hat{R}]} < R < \hat{R} + K \sqrt{\hat{ECM}[\hat{R}]}\ ] = 1 - \alpha = 1 - 1/K^{2}$$

II) Si $CV(\hat{\bar{X}}st) > 0.1$, no se considera el sesgo despreciable. Por lo tanto, no se puede calcular el intervalo de confianza.

Hay que aumentar el tamaño de muestra.

2.5.3. DETERMINACION DEL TAMAÑO DE MUESTRA

Se presentan los procedimientos para determinar el tamaño de muestra utilizando diferentes asignaciones. Antes de describirlos, se presentan los diferentes tipos de distribución de la muestra total, a los estratos.

2.5.3.1. ASIGNACION DE LA MUESTRA A LOS ESTRATOS

Después de determinar el tamaño de muestra (n), existen varias formas de distribuirlo a los diferentes estratos. Aquí se presentan tres tipos:

1) Asignación proporcional
2) Asignación óptima
3) Asignación de Neyman

1) Asignación proporcional

El tamaño de muestra (n) se distribuye de manera proporcional a los tamaños de los estratos (Nh). Se cumple la siguiente relación:

$$\frac{nh}{n} = \frac{Nh}{N}$$

Este tipo de asignación se usa cuando:

a) no se conocen las varianzas dentro de los estratos (Syh^2), o bien no se tienen buenos estimadores de dichas varianzas.

b) la varianza y el costo unitario dentro de cada estrato son muy parecidos entre estratos.

c) se quiere obtener información sobre muchas variables.

El tamaño de muestra por estrato se obtiene de la siguiente manera:

$$nh = \left(\frac{Nh}{N}\right)n$$

2) Asignación óptima

Este tipo de asignación se utiliza cuando se pretende obtener una distribución que minimice la varianza de un estimador para un costo total fijo o que minimice el costo total para una varianza del estimador fija.

Supóngase que el costo total de la investigación puede describirse de la siguiente manera:

$$Ct = C1 + \sum_{h=1}^{L} nh \; Ch$$

Ct: Costo total
C1: Costo fijo o administrativo
Ch: Costo unitario de obtención de información en el h-ésimo estrato

Entonces la distribución que minimiza el costo producto por varianza del estimador (es la misma ya sea Y o Y) es

$$nh = \left(\frac{Nh \; Syh \, / \sqrt{Ch}}{\sum_{h=1}^{L} Nh \; Syh \, / \sqrt{Ch}} \right) n$$

(Ver Cochran 1974 "Técnicas de Muestreo" Página 138)

Obesérvese que en esta distribución influyen tres factores:

a) Tamaño de estrato
b) Varianza dentro de estrato
c) Costo por unidad de cada estrato

Este tipo de distribución se utiliza cuando las varianzas y los costos varían entre estratos y se tiene buena información acerca de éllas. No conviene utilizarla cuando se requiere de información de muchas variables simultáneamente.

3) Asignación de Neyman

Este tipo de asignación se deriva de la asignación óptima cuando se puede considerar que los costos por unidad son iguales entre estratos. Se tendrá entonces la siguiente relación:

$$Ct = C1 + n \, C'$$

C': Costo por unidad

La muestra total se distribuye a los estratos de la siguiente manera:

$$nh = \left(\frac{Nh\ Syh}{\sum_{h=1}^{L} Nh\ Syh} \right) n$$

2.5.3.2. PROCEDIMIENTOS PARA LA DETERMINACION DEL TAMAÑO DE MUESTRA TOTAL

En esta parte se presentan los procedimientos para determinar el tamaño de muestra cuando se pretende estimar los siguientes parámetros:

1) Media
2) Total
3) Proporción

utilizando las siguientes asignaciones:

i) Asignación óptima
ii) Asignación de Neyman
iii) Asignación proporcional

En el caso en el que se utilice la asignación óptima y de Neyman se proporcionan dos alternativas:

a) Determinar el tamaño de muestra que minimice la varianza para un costo total fijo

b) Determinar el tamaño de muestra que minimice el costo total para una varianza fija

En el caso de la asignación proporcional se proporcionan las siguientes alternativas:

a) Determinar el tamaño de muestra que se requiera para tener una precisión y confianza prefijadas

b) Determinar el tamaño de muestra que se requiera para obtener un coeficiente de variación prefijado

1) Determinación del tamaño de muestra para estimar una media poblacional

i) Si se piensa utilizar una asignación óptima de la muestra a los estratos, el tamaño de muestra se determina según dos alternativas:

a) Si se quiere minimizar la varianza para un costo total fijo(CO)

$$n = \frac{(CO - C1)\left(\sum_{h=1}^{L} Nh\ Syh / \sqrt{Ch}\right)}{\sum_{h=1}^{L} Nh\ Syh \sqrt{Ch}}$$

y entonces $$nh = \left(\frac{Nh\ Syh / \sqrt{Ch}}{\sum_{h=1}^{L} Nh\ Syh / \sqrt{Ch}}\right) n$$

b) Si se quiere minimizar el costo total para una varianza fija(VO)

$$n = \frac{\left(\sum_{h=1}^{L} Nh\ Syh\sqrt{Ch}\right)\left(\sum_{h=1}^{L} Nh\ Syh / \sqrt{Ch}\right)}{N^2\ Vo + \sum_{h=1}^{L} Nh\ Syh^2}$$

y entonces $$nh = \left(\frac{Nh\ Syh / \sqrt{Ch}}{\sum_{h=1}^{L} Nh\ Syh / \sqrt{Ch}}\right) n$$

ii) Si se piensa utilizar una asignación de Neyman de la muestra a los estratos, el tamaño de muestra se determina según dos alternativas:

a) Si se quiere minimizar la varianza para un costo total fijo(CO)

$$n = \frac{CO - C1}{C'} \quad \text{y entonces} \quad nh = \left(\frac{Nh\ Syh}{\sum_{h=1}^{L} Nh\ Syh}\right) n$$

b) Si se quiere minimizar el costo total para una varianza fija(VO)

$$n = \frac{\left(\sum_{h=1}^{L} Nh\ Syh\right)^2}{N^2\ VO + \sum_{h=1}^{L} Nh\ Syh^2} \quad \text{y entonces} \quad nh = \left(\frac{Nh\ Syh}{\sum_{h=1}^{L} Nh\ Syh}\right) n$$

iii) Si se piensa utilizar una asignación proporcional de la muestra a los estratos, el tamaño de muestra se determina según dos alternativas:

a) Si se quiere fijar la precisión (BO) y la confianza (1-α)

$$n' = \frac{\left(\sum_{h=1}^{L} Wh\ Syh^{2}\right) Z^{2}_{\alpha/2}}{BO^{2}} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

y entonces $nh = \left(\frac{Nh}{N}\right) n$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{\sum_{h=1}^{L} Wh\ Syh^{2}}{CVO^{2}\ \bar{Y}^{2}} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

y entonces $nh = \left(\frac{Nh}{N}\right) n$

2) Determinación del tamaño de muestra para estimar un total poblacional

i) Asignación óptima

a) Si se quiere minimizar la varianza para un costo total fijo(CO)

$$n = \frac{(CO - C1)\left(\sum_{h=1}^{L} Nh\ Syh / \sqrt{Ch}\right)}{\sum_{h=1}^{L} Nh\ Syh \sqrt{Ch}}$$

y entonces $nh = \left(\frac{Nh\ Syh / \sqrt{Ch}}{\sum_{h=1}^{L} Nh\ Syh / \sqrt{Ch}}\right) n$

b) Si se quiere minimizar el costo total para una varianza fija(VO)

$$n = \frac{\left(\sum_{h=1}^{L} Nh\ Syh\sqrt{Ch}\right)\left(\sum_{h=1}^{L} Nh\ Syh\ /\sqrt{Ch}\right)}{VO + \sum_{h=1}^{L} Nh\ Syh^{2}}$$

$$\text{y entonces } nh = \left(\frac{Nh\ Syh\ /\sqrt{Ch}}{\sum_{h=1}^{L} Nh\ Syh\ /\sqrt{Ch}}\right) n$$

ii) Asignación de Neyman

a) Si se quiere minimizar la varianza para un costo total fijo(CO)

$$n = \frac{CO - C1}{C'} \quad \text{y entonces } nh = \left(\frac{Nh\ Syh}{\sum_{h=1}^{L} Nh\ Syh}\right) n$$

b) Si se quiere minimizar el costo total para una varianza fija(VO)

$$n = \frac{\left(\sum_{h=1}^{L} Nh\ Syh\right)^{2}}{VO + \sum_{h=1}^{L} Nh\ Syh^{2}} \quad \text{y entonces} \quad nh = \left(\frac{Nh\ Syh}{\sum_{h=1}^{L} Nh\ Syh}\right) n$$

iii) Asignación proporcional

a) Si se quiere fijar la precisión (BO) y la confianza (1-α)

$$n' = \frac{N^{2}\left(\sum_{h=1}^{L} Wh\ Syh^{2}\right) Z^{2}_{\alpha/2}}{BO^{2}} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

$$\text{y entonces} \quad nh = \left(\frac{Nh}{N}\right) n$$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{\sum_{h=1}^{L} Wh\ Syh^2}{CVO^2\ \bar{Y}^2} \qquad n = \frac{1}{\frac{1}{n'} + \frac{1}{N}}$$

y entonces $nh = \left(\frac{Nh}{N}\right)n$

3) Determinación del tamaño de muestra para estimar una proporción poblacional

i) Asignación óptima

a) Si se quiere minimizar la varianza para un costo total fijo(CO)

$$n = \frac{(CO - C1)\left(\sum_{h=1}^{L} Nh\sqrt{\frac{Nh\ Ph\ Qh}{(Nh-1)Ch}}\right)}{\sum_{h=1}^{L} Nh\sqrt{\frac{Nh\ Ph\ Qh\ Ch}{Nh - 1}}}$$

y entonces $$nh = \left(\frac{Nh\sqrt{\frac{Nh\ Ph\ Qh}{(Nh-1)Ch}}}{\sum_{h=1}^{L} Nh\sqrt{\frac{Nh\ Ph\ Qh}{(Nh-1)Ch}}}\right)n$$

b) Si se quiere minimizar el costo total para una varianza fija(VO)

$$n = \frac{\left(\sum_{h=1}^{L} Nh\sqrt{\frac{Nh\ Ph\ Qh\ Ch}{Nh - 1}}\right)\left(\sum_{h=1}^{L} Nh\sqrt{\frac{Nh\ Ph\ Qh}{(Nh-1)Ch}}\right)}{N^2\ VO + \sum_{h=1}^{L} Nh\frac{Nh\ Ph\ Qh}{Nh - 1}}$$

y entonces $$nh = \left(\frac{Nh\sqrt{\frac{Nh\ Ph\ Qh}{(Nh-1)Ch}}}{\sum_{h=1}^{L} Nh\sqrt{\frac{Nh\ Ph\ Qh}{(Nh-1)Ch}}}\right)n$$

ii) Asignación de Neyman

a) Si se quiere minimizar la varianza para un costo total fijo(CO)

$$n = \frac{CO - C1}{C'} \quad \text{y entonces } nh = \left(\frac{Nh \sqrt{\dfrac{Nh\ Ph\ Qh}{Nh - 1}}}{\sum_{h=1}^{L} Nh \sqrt{\dfrac{Nh\ Ph\ Qh}{Nh - 1}}} \right) n$$

b) Si se quiere minimizar el costo total para una varianza fija(VO)

$$n = \frac{\left(\sum_{h=1}^{L} Nh \sqrt{\dfrac{Nh\ Ph\ Qh}{Nh - 1}} \right)^2}{N^2\ VO + \sum_{h=1}^{L} Nh \dfrac{Nh\ Ph\ Qh}{Nh - 1}}$$

$$\text{y entonces } nh = \left(\frac{Nh \sqrt{\dfrac{Nh\ Ph\ Qh}{Nh - 1}}}{\sum_{h=1}^{L} Nh \sqrt{\dfrac{Nh\ Ph\ Qh}{Nh - 1}}} \right) n$$

iii) Asignación proporcional

a) Si se quiere fijar la precisión (BO) y la confianza $(1-\alpha)$

$$n' = \frac{\left(\sum_{h=1}^{L} Wh \dfrac{Nh\ Ph\ Qh}{Nh - 1} \right) Z^2_{\alpha/2}}{BO^2} \qquad n = \frac{1}{\dfrac{1}{n'} + \dfrac{1}{N}}$$

$$\text{y entonces } nh = \left(\frac{Nh}{N} \right) n$$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n' = \frac{\sum_{h=1}^{L} Wh \dfrac{Nh\ Ph\ Qh}{Nh - 1}}{CVO^2\ P^2} \qquad n = \frac{1}{\dfrac{1}{n'} + \dfrac{1}{N}}$$

$$\text{y entonces } nh = \left(\frac{Nh}{N} \right) n$$

2.6. MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMAÑO CON REEMPLAZO

En esta sección se describen los procedimientos de estimación de una media y un total poblacionales y de determinación del tamaño de muestra utilizando la misma notación que en el muestreo aleatorio simple.

2.6.1. PROCEDIMIENTOS DE ESTIMACION

1) Media

Estimador

$$\hat{\bar{Y}} = \frac{1}{N}\,\frac{\sum_{i=1}^{n} y_i/p_i}{n} \qquad p_i = \frac{X_i}{X}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V(\hat{\bar{Y}}) = \frac{\sum_{i=1}^{N} p_i \left(Y_i/p_i - Y \right)^2}{N^2\, n}$$

Estimador de la varianza del estimador

$$\hat{V}(\hat{\bar{Y}}) = \frac{\sum_{i=1}^{n} \left(y_i/p_i - \frac{1}{n} \sum_{i=1}^{n} y_i/p_i \right)^2}{N^2\, (n-1)n}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

a) Si el tamaño de muestra es grande ($n \geq 30$), se aproxima a la normal.

$$P\left[\hat{\bar{Y}} - Z_{\alpha/2} \sqrt{\hat{V}(\hat{\bar{Y}})} < \bar{Y} < \hat{\bar{Y}} + Z_{\alpha/2} \sqrt{\hat{V}(\hat{\bar{Y}})} \right] = 1 - \alpha$$

b) Si el tamaño de muestra es pequeño ($n < 30$), se usa la desigualdad de Tchebycheff.

$$P\left[\hat{\bar{Y}} - K \sqrt{\hat{V}(\hat{\bar{Y}})} < \bar{Y} < \hat{\bar{Y}} + K \sqrt{\hat{V}(\hat{\bar{Y}})} \right] = 1 - \alpha = 1 - 1/K^2$$

2) Total

Estimador

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} \qquad p_i = \frac{X_i}{X}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V[\hat{Y}] = \frac{1}{n} \sum_{i=1}^{N} p_i \left(Y_i/p_i - Y \right)^2$$

Estimador de la varianza del estimador

$$\hat{V}[\hat{Y}] = \frac{\sum_{i=1}^{n} \left(y_i/p_i - \frac{1}{n} \sum_{i=1}^{n} y_i/p_i \right)^2}{n(n-1)}$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

a) Si el tamaño de muestra es grande ($n \geq 30$), se aproxima a la normal.

$$P[\hat{Y} - Z_{\alpha/2} \sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + Z_{\alpha/2} \sqrt{\hat{V}[\hat{Y}]}\,] = 1 - \alpha$$

b) Si el tamaño de muestra es pequeño ($n<30$), se usa la desigualdad de Tchebycheff.

$$P[\hat{Y} - K \sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + K \sqrt{\hat{V}[\hat{Y}]}\,] = 1 - \alpha = 1 - 1/K^2$$

2.6.2. DETERMINACION DEL TAMAÑO DE MUESTRA

En esta parte, se presentan los procedimientos para la determinación del tamaño de muestra para estimar una media y un total, proporcionando dos alternativas:

a) Determinar el tamaño de muestra que se requiera para tener una precisión y confianza prefijadas

b) Determinar el tamaño de muestra que se requiera para obtener un coeficiente de variación prefijado

1) Determinación del tamaño de muestra para estimar una media poblacional

a) Si se quiere fijar la precisión (BO) y la confianza (1-α)

$$n = \frac{Z^2_{\alpha/2} \sum_{i=1}^{N} pi \left(Yi/pi - Y \right)^2}{N^2 \, BO^2}$$

Nota: El valor de $\sum_{i=1}^{N} pi \left(Yi/pi - Y\right)^2$ se puede estimar a partir de una muestra piloto mediante

$$\frac{\sum_{i=1}^{n} \left(yi/pi - \frac{1}{n} \sum_{i=1}^{n} yi/pi \right)^2}{n - 1}$$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n = \frac{\sum_{i=1}^{N} pi \left(Yi/pi - Y \right)^2}{N^2 \bar{Y}^2 \, CVO^2}$$

2) Determinación del tamaño de muestra para estimar un total poblacional

a) Si se quiere fijar la precisión (BO) y la confianza (1-α)

$$n = \frac{Z^2_{\alpha/2} \sum_{i=1}^{N} pi \left(Yi/pi - Y \right)^2}{BO^2}$$

b) Si se quiere fijar el coeficiente de variación (CVO)

$$n = \frac{\sum_{i=1}^{N} pi \left(Yi/pi - Y \right)^2}{N^2 \, CVO^2 \, \bar{Y}^2}$$

2.7. MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON PROBABILIDAD PROPORCIONAL AL TAMAÑO CON REEMPLAZO Y EN LA SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

Se considerará que el valor de la variable auxiliar Xi para la selección con probabilidad proporcional al tamaño, es el número de unidades secundarias dentro de la i-ésima unidad primaria.

En este esquema de muestreo, se proporcionan solamente los procedimientos de estimación de una media y un total.

2.7.1. NOTACION

N : Total de unidades primarias en la población

n : Total de unidades primarias en la muestra

M : Total de unidades secundarias en la población

$$M = \sum_{i=1}^{N} Mi$$

Mi: Total de unidades secundarias en la i-ésima unidad primaria

mi: Total de unidades secundarias muestreadas en la i-ésima unidad primaria

Yij: Valor de la característica de interés de la j-ésima unidad secundaria de la i-ésima unidad primaria

Yi: Total poblacional de la característica de interés en la i-ésima unidad primaria

$$Yi = \sum_{j=1}^{Mi} Yij$$

$\bar{Y}$: Media poblacional de la característica de interés

$$\bar{Y} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{Mi} Yij}{M}$$

Y : Total poblacional de la característica de interés

$$Y = \sum_{i=1}^{N} \sum_{j=1}^{Mi} Yij$$

S_{wi}^2: Varianza de la característica de interés dentro de la i-ésima unidad primaria.

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

y_{ij}: Valor de la característica de interés para la j-ésima unidad secundaria de la muestra en la i-ésima unidad primaria

$\bar{y}_i$: Media muestral de la característica de interés en la i-ésima unidad primaria.

$$\bar{y}_i = \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i}$$

Se usa el signo ^ para estimadores.

2.7.2. PROCEDIMIENTOS DE ESTIMACION

1) Media

Estimador

$$\hat{\bar{Y}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V(\hat{\bar{Y}}) = \left[\frac{1}{nN^2} \sum_{i=1}^{N} p_i \left(Y_i/p_i - Y \right)^2 + \frac{1}{nN^2} \sum_{i=1}^{N} \frac{M_i^2}{p_i} \left(1 - \frac{m_i}{M_i} \right) \frac{S_{wi}^2}{m_i} \right]$$

donde $p_i = \dfrac{M_i}{M}$

Estimador de la varianza del estimador

$$\hat{V}[\hat{\bar{Y}}] = \frac{1}{M^2 n(n-1)} \sum_{i=1}^{n} \left(\frac{\hat{Y}_i}{p_i} - \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{Y}_i}{p_i} \right)^2 \quad \text{donde } p_i = \frac{M_i}{M} \quad \text{donde } \hat{Y}_i = M_i \hat{\bar{Y}}_i$$

Intervalos de confianza al $(1-\alpha)\times100\%$

a) Si la variable (y) se distribuye normalmente o el total de unidades primarias en la muestra es grande ($n\geq30$) o el total de unidades secundarias muestreadas en todas las unidades primarias es grande ($m_i\geq30$, para toda i), se aproxima a la normal.

$$P[\hat{\bar{Y}} - Z_{\alpha/2} \sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + Z_{\alpha/2} \sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1 - \alpha$$

b) Si no se cumple ninguna de las condiciones mencionada en a), se usa la desigualdad Tchebycheff.

$$P[\hat{\bar{Y}} - K \sqrt{\hat{V}[\hat{\bar{Y}}]} < \bar{Y} < \hat{\bar{Y}} + K \sqrt{\hat{V}[\hat{\bar{Y}}]}] = 1 - \alpha = 1 - 1/K^2$$

2) Total

Estimador

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i}$$

Nota: Este es un estimador insesgado.

Varianza del estimador

$$V[\hat{Y}] = \left[\frac{1}{n} \sum_{i=1}^{N} p_i \left(\frac{Y_i}{p_i} - Y \right)^2 + \frac{1}{n} \sum_{i=1}^{N} \frac{M_i^2}{p_i} \left(1 - \frac{m_i}{M_i} \right) \frac{S_{wi}^2}{m_i} \right]$$

donde $p_i = \frac{M_i}{N}$

Estimador de la varianza del estimador

$$\hat{V}[\hat{Y}] = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{\hat{Y}_i}{p_i} - \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{Y}_i}{p_i} \right)^2 \qquad \text{donde } p_i = \frac{M_i}{M}$$

$$\text{donde } \hat{Y}_i = M_i \bar{\hat{Y}}_i$$

Intervalos de confianza al $(1-\alpha)\times 100\%$

a) Si la variable (y) se distribuye normalmente o el total de unidades primarias en la muestra es grande ($n \geq 30$) o el total de unidades secundarias muestreadas es grande en todas las unidades primarias en la muestra ($m_i \geq 30$, para toda i), se aproxima a la normal.

$$P[\hat{Y} - Z_{\alpha/2} \sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + Z_{\alpha/2} \sqrt{\hat{V}[\hat{Y}]}\,] = 1 - \alpha$$

b) Si no se cumple ninguna de las condiciones mencionadas en a), se usa la desigualdad de Tchebycheff.

$$P[\hat{Y} - K \sqrt{\hat{V}[\hat{Y}]} < Y < \hat{Y} + K \sqrt{\hat{V}[\hat{Y}]}\,] = 1 - \alpha = 1 - 1/K^2$$

3. MANUAL DE OPERACION

3.1. INTRODUCCION

En este trabajo se proporcionan tres módulos de PASCAL para cuatro tipos de diseño de muestra.

1) Muestreo aleatorio simple sin reemplazo
2) Muestreo estratificado aleatorio
3) Muestreo con probabilidad proporcional al tamaño con reemplazo
4) Muestreo bietápico con selección en la primera etapa con probabilidad proporcional al número de unidades secundarias con reemplazo y en la segunda etapa con muestreo aleatorio simple sin reemplazo

1) Para muestreo aleatorio simple sin reemplazo se disponen de procedimientos para estimar los siguientes parámetros:

a) Media
b) Total
c) Proporción
d) Razón

Estos procedimientos permiten obtener estimadores simples de todos los parámetros y, para la media y el total, permiten además obtener estimadores de razón y regresión.

También se disponen de procedimientos para determinar el tamaño de muestra para cada caso y dependiendo del estimador que se elija.

2) Para muestreo estratificado aleatorio se disponen de procedimientos para estimar los siguientes parámetros:

a) Media
b) Total
c) Proporción
d) Razón

Estos procedimientos permiten obtener estimadores simples de todos los parámetros y para la media y el total, permiten además obtener estimadores de razón y regresión.

Para este tipo de diseño se presentan además dos opciones para la obtención de estimadores de razón y regresión: estimadores combinados y estimadores separados.

Unicamente se proporcionan procedimientos para determinar tamaño de muestra cuando se utilizan estimadores simples de los siguientes parámetros:

a) Media

b) Total
c) Proporción

No se incluyen procedimientos de determinación de tamaño de muestra cuando se utilizan estimadores de razón y regresión.

3) Para muestreo con probabilidad proporcional al tamaño con reemplazo se disponen de procedimientos de estimación y determinación de tamaño de muestra para los siguientes parámetros:

a) Media
b) Total

4) Para muestreo bietápico con selección con probabilidad proporcional al tamaño con reemplazo en la primera etapa y selección aleatoria simple sin reemplazo en la segunda etapa, se disponen de procedimientos para estimar los siguientes parámetros:

a) Media
b) Total

Para este tipo de diseño no se incluyen procedimientos para determinación del tamaño de muestra.

3.2. ?COMO INICIAR?

Las microcomputadoras (PRINTAFORM) que están disponibles en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas(I.I.M.A.S) tienen una pantalla, un teclado y dos unidades de disco. Se dispone de una impresora. Para distinguir a las dos unidades de disco, se denomina a la de arriba "UNIDAD-A" y a la de abajo "UNIDAD-B".

Para poder utilizar este paquete se necesita un disco que contiene los siguientes módulos y archivos:

COMMAND.COM : sistema operativo de MS-DOS(Version 2.0).
AUTOEXEC.BAT : archivo para correr directamente los programas de este paquete.
MUESTREO.COM : módulo para iniciar.
SIMPLE.COM y
SIMPLE.000 : módulo para el muestreo aleatorio simple sin reemplazo
ESTRATIF.COM y
ESTRATIF.000 : módulo para el muestreo estratificado aleatorio
PPT.COM : modulo para el muestreo con probabilidad proporcional al tamaño y el muestreo bietápico con selección con probabilidad proporcional al tamaño con reemplazo en la primera etapa y selección aleatoria simple sin reemplazo en la segunda etapa.
ZTABLE.DTA : archivo que contiene los valores de las tablas de la normal estandar.

tTABLE.DTA : archivo que contiene los valores de las tablas de la t de Student.

Este disco deberá ser colocado en la UNIDAD-A.

En el caso en el que los datos de la muestra (valores de la característica de interés y/o variable auxiliar) se quieran leer y/o grabar en un archivo, se necesita otro disco, el cual se coloca en la UNIDAD-B. Si este disco es nuevo, se deberá formatear antes de utilizarlo usando el comando FORMAT.(Ver el manual de MS-DOS).

Después de colocar ambos discos en las unidades, se enciende la microcomputadora. El sistema operativo y el programa para iniciar se cargan en la memoria automáticamente y aparece la siguiente imagen en pantalla:

(I-1)

```
**     ** *     * ****** ***** ***** ****** ***** *****
* *   * * *     * *      *       *    *    * *     *   *
*  ** * * *     * *****  *****   *    ****** ****  *   *
*       * *     *            *   *    *   *  *     *   *
*       * ****** ****** *****    *    *    * ***** *****

1-MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

2-MUESTREO ESTRATIFICADO ALEATORIO

3-MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO
                              O
  MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON PRO-
  BABILIDAD PROPORCIONAL AL TAMANO  CON REEMPLAZO Y EN LA SEGUNDA
  ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

!Numero que corresponde!
_
```

Se debe teclear un número entre 1 y 3 que corresponde al diseño de muestra de interés y posteriormente teclear 'RETURN'. Después de introducirlo, aparece otra imagen en pantalla dependiendo del diseño de muestra que se haya escogido.

1) MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

```
*****************************************
* MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO *
*****************************************
1-Obtener un estimador?
2-Determinar un tamano de muestra?
3-Ambos?
!Numero que corresponde!
_
```

(I-2)

2) MUESTREO ESTRATIFICADO ALEATORIO

```
**********************************
* MUESTREO ESTRATIFICADO ALEATORIO *
**********************************
1-Obtener un estimador?
2-Determinar un tamano de muestra?
3-Ambos?
|Numero que corresponde|
_
```

(I-3)

3) MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMAÑO CON REEMPLAZO o MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO y EN LA SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

(I-4)

```
1-MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO
2-MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON PROBABI-
  LIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO y EN LA SEGUNDA ETAPA
  CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO
|Numero que corresponde|
_
```

En este módulo al escoger alguno de los dos números, aparece para el diseño de muestra la siguiente imagen en pantalla:

3-1) MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMAÑO CON REEMPLAZO

(I-5)

```
****************************************************************
* MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO*
****************************************************************
1-Obtener un estimador?
2-Determinar un tamano de muestra?
3-Ambos?
|Numero que corresponde|
_
```

3-2) MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON PROBABILIDAD PROPORCIONAL AL TAMAÑO CON REEMPLAZO y EN LA SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO

(I-6)

```
**************************************************************
* MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA       *
* CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO      *
* y EN LA SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE        *
* SIN REEMPLAZO                                              *
**************************************************************
?Parametro a estimar?
1-Media
2-Total
!Numero que corresponde!
_
```

Los módulos de este paquete son interactivos y fáciles de manejar. Contestando a cada pregunta se obtienen los resultados deseados.

Nota: Cuando se contesta a una pregunta que aparece en pantalla, primero se debe teclear lo que se necesita y posteriormente teclear RETURN.

De aquí en adelante, para distinguir a los diferentes diseños de muestra, se utilizarán las siguientes denominaciones:

M.A.S : Muestreo aleatorio simple sin reemplazo
ESTRATIFICADO : Muestreo estratificado aleatorio
P.P.T : Muestreo con probabilidad proporcional al tamaño con reemplazo
BIETAPICO : Muestreo bietápico con selección con probabilidad proporcional al tamaño con reemplazo en la primera etapa y selección aleatoria simple sin reemplazo en la segunda etapa

3.3. FUNCIONAMIENTO DE LOS MODULOS DE ESTE PAQUETE

Como se ve en I-2, I-3 y I-5 de 3.2, cada módulo de este paquete excepto el que corresponde al BIETAPICO proporciona tres alternativas:

1) Estimar un parámetro

2) Determinar un tamaño de muestra

3) Estimar un parámetro y determinar un tamaño de muestra

Para seleccionar una de las tres alternativas, se deben tener las siguientes consideraciones:

a) Si se tienen los datos vírgenes de una muestra piloto, se debe escoger [3], en cuyo caso, se calcula la información que se necesita para la determinación de tamaño de muestra a través de los procedimientos de la sección de estimación.

b) Si se tiene una muestra piloto pero la información que se requiere para determinar el tamaño de muestra ya está calculada, se deberá elegir [2].

c) Si se tiene un problema para estimar un parámetro y un problema para determinar un tamaño de muestra, pero las características de interés son distintas, se debe escoger [1] o [2] por separado y tratarlos como problemas diferentes corriendo dos veces el módulo.

Para BIETAPICO se proporcionan solamente los procedimientos para estimar un parámetro. Este paquete no incluye los procedimientos para la determinación de tamaño de muestra en este caso.

Cada vez que termine la ejecución de una de las tres alternativas se pregunta:

```
?Quiere empezar de nuevo? (SI o NO)
_
```

(I-7)

al contestar 'SI', se puede repetir tantas veces como uno desee el módulo elegido.

Además, los procedimientos para determinación del tamaño de muestra también se pueden ejecutar varias veces respondiendo 'SI' a la pregunta:

```
?Quiere determinar otro tamano de muestra?
(SI o NO)
_
```

(I-8)

Cuando se quiera terminar la ejecución, se deberá contestar 'NO' a las preguntas de arriba, y aparece la siguiente imagen en pantalla:

```
************************
* Terminamos !GRACIAS! *
************************
```

(I-9)

y se retiran los discos de las unidades y se apaga la microcomputadora.

3.4. VERIFICACION y CORRECCION DE LOS DATOS y CANCELACION DE MODULOS

3.4.1. VERIFICACION

Los programas de este trabajo verifican lógicamente algunos datos que se han introducido.

1) Para cualquier tipo de diseño de muestra el tamaño de ésta no debe exceder al de la población y no debe ser 0.

2) Cuando se estima una proporción, el total de unidades que poseen la característica de interés no debe exceder al tamaño de muestra ni ser 0.

3) El total de la variable auxiliar para los elementos de la muestra no debe exceder al total poblacional de la misma.

4) En muestreo estratificado aleatorio, para la determinación del tamaño de muestra que minimice la varianza del estimador para un costo total prefijado, este no deberá ser menor que la suma de los costos constantes más el costo de obtencion de informacion.

5) El total de unidades en la población así como los totales de variable auxiliar y de otros parámetros para determinar el tamaño de muestra no deben ser 0.

Se toma el caso del tamaño de muestra de M.A.S como un ejemplo:

El programa pide el tamaño de muestra de la forma siguiente:

```
?Tamano de muestra (n)?
_
```

(I-10)

Se deberá introducir un número entero para el tamaño de muestra que se tiene. El tamaño de muestra debe estar entre el tamaño de la población y cero. Si no esta dentro de este rango, aparecerá el siguiente mensaje de error en pantalla:

```
@Error(No debe ser mayor que el tamano de poblacion)
      (No debe ser 0)
!Introduzca otra vez!
_
```

(I-11)

Se debe introducir el numero correcto nuevamente.

De este modo, todos los programas de este trabajo verifican en la medida de la posible la información y mandan el mensaje de error pidiendo la información correcta.

3.4.2. CORRECCION DE ERRORES DENTRO DE MODULOS

Este paquete únicamente permite la corrección (sin necesidad de empezar algún proceso de nuevo) de la información requerida por el mismo, para los casos en que ésta sea cuantiosa, por ejemplo la información relativa a la característica de interés (y), la relativa a la variable auxiliar (x), total de unidades por estrato, tamaño de muestra por estrato etc. Se proporciona la oportunidad de revisarlos y modificarlos después de haberlos introducido o leído de un archivo, contestando un "SI" a las preguntas:

```
?Quiere ver los datos? (SI o NO)
_
```

(I-12)

```
?Quiere modificar algun dato? (SI o NO)
_
```

(I-13)

En el caso de M.A.S y de P.P.T, la variable y[i] que aparece en pantalla representa el valor de la característica de interés para la i-ésima observación de la muestra y la variable x[i] representa el valor de la variable auxiliar para la i-ésima observación de la muestra.

En el caso de ESTRATIFICADO, la variable y[h,i] que aparece en pantalla representa el valor de la característica de interés para la i-ésima observación del h-ésimo estrato y la variable x[h,i] representa el valor de la variable auxiliar para la i-ésima observación del h-ésimo estrato.

En el caso de BIETAPICO, la variable y[i,j] que aparece en pantalla representa el valor de la característica de interés para la j-ésima observación de la i-ésima unidad primaria de la muestra.

La parte de modificación se pueden repetir varias veces respondiendo 'SI' a la pregunta:

```
?Quiere modificar otro dato? (SI o NO)
_
```

(I-14)

3.4.3. CANCELACION DE MODULOS

En algunas ocasiones, se puede querer salir del módulo en curso por distintas razones, por ejemplo se introdujo una información erroneamente y no se puede corregir o se eligió un módulo inadecuado etc. En estos casos, se deberá presionar [Ctrl] y teclear [C] . [Ctrl] y [C] significa que se cancela el módulo. Después de teclear, aparece "A>" en pantalla.

Si se quiere volver a elegir un módulo, se deberá teclear una de las siguientes palabras dependiendo del diseño de muestra de interés:

A>SIMPLE : para M.A.S
A>ESTRATIF : para ESTRATIFICADO
A>PPT : para P.P.T y BIETAPICO

y aparece una de las imagenes denotadas anteriormente como I-2, I-3 o I-4 en 3.2 dependiendo del diseño elegido.

Si se quiere terminar, se retiran los discos de las unidades y se apaga la microcomputadora.

Nota: Se puede cancelar el programa cuando se quiera tecleando [Ctrl] [C] al mismo tiempo.
Cuando aparece un mensaje de error y "A>" en pantalla por algún problema, se puede elegir de nuevo un módulo mediante el procedimiento anteriormente mencionado.

3.5. LECTURA E IMPRESION DE DATOS Y RESULTADOS

3.5.1. LEER UN ARCHIVO

Si se tienen los valores de la variable (y) y la variable (x) o los totales de unidades por estratos que poseen la característica de interés en un archivo, el paquete puede leer directamente de dicho archivo.

En primer lugar, se debe revisar que el disco en el cual se encuentra el archivo que contiene la información, esté colocado en la UNIDAD-B.

En segundo lugar, se debe contestar 'SI' a la pregunta:

```
?Tiene los datos en un archivo? (SI o NO)
_
```

(I-15)

En tercer lugar, se debe introducir el nombre del archivo cuando aparece el siguiente mensaje en pantalla:

```
?Nombre de archivo = b:_
```

(I-16)

Se debe escribir el nombre con un máximo de 12 letras de la siguiente forma:

XXXXXXXX.AAA

XXXXXXXX es el nombre principal que no debe exceder a 8 letras, y el cual debe necesariamente ir seguido por un punto.
AAA es el nombre de extensión que debe constar exactamente de 3 letras.

El programa agrega una "b:" al nombre del archivo que se haya introducido, y lo busca en el disco de la UNIDAD-B. Si no lo encuentra, aparece el siguiente mensaje de error en pantalla:

```
@Error(No se encuentra el archivo = XXXXXXXX.AAA)
!Introduzca otra vez!
?Nombre de archivo = b:_
```

(I-17)

Estructura del archivo

1) La única información que puede introducirse a través de un archivo es: los valores de la variable (y), los valores de la variable (x) (en caso necesario) o los totales de unidades por estrato que poseen la característica de interés cuando se desee estimar una proporción en ESTRATIFICADO.

2) Cada registro debe contener un valor de la variable (y) y un valor de la variable (x) (en caso necesario) guardando este orden. Cuando se estima una proporción en ESTRATIFICADO, cada registro deberá contener el total de unidades por estrato que poseen la característica de interés. (Format libre)

3) El archivo debe tener un número igual o mayor de registros que el tamaño de muestra o en el caso de estimar una proporción en ESTRATIFICADO debe tener un número igual o mayor de registros que el número de estratos.

   Si el archivo no lo cumple, se cancela el módulo y aparece "A>" en pantalla.

4) En el caso de ESTRATIFICADO, los valores deben introducirse sucesivamente por estrato, es decir, primero todos los del primer estrato, después todos los del segundo estrato, y así sucesivamente.

5) En el caso de BIETAPICO, los valores deben introducirse sucesivamente por unidad primaria, es decir, primero todos los de la primera unidad primaria, después todos los de la segunda unidad primaria, y así sucesivamente

3.5.2. ESCRIBIR UN ARCHIVO

Una vez que se tengan los datos de la variable (y) y la variable (x) o los totales de unidades por estrato que poseen la característica de interés cargados en la memoria y ya corregidos, si se desea guardarlos en un archivo, éste se grabará en el disco que se encuentra colocado en la UNIDAD-B, de la siguiente manera:

En primer lugar, se debe revisar que el disco en el cual se quiere grabar el archivo, esté colocado en la UNIDAD-B.

En segundo lugar, se deben contestar 'SI' a la pregunta:

```
?Quiere guardar los datos en un archivo? (SI o NO)
_
```
(I-18)

En tercer lugar, se debe introducir el nombre del archivo cuando aparece el siguiente mensaje en pantalla:

```
?Nombre de archivo = b:_
```
(I-19)

Se debe escribir el nombre de archivo con un máximo de 12 letras de la siguiente forma:

XXXXXXXX.AAA   con las especificaciones mencionadas en I-16 de 3.5.1

Si no se tiene ninguna idea para poner un nombre AAA, se recomienda poner DTA.

?Como se graba la información en el archivo?

La estructura del archivo es de la misma forma que se describió en 3.5.1.

Nota: Si se eligió el mismo nombre del archivo que con el que se leyó, los datos se graban encima.

3.5.3. IMPRESION

La información que puede imprimirse es la siguiente:

1) los valores de la característica de interés (yi) para cada elemento de la muestra cuando se estima cualquier parámetro excepto una proporción.

2) los valores de la variable auxiliar (xi) para cada elemento de la muestra

3) los totales de unidades por estrato que poseen la característica de interés en ESTRATIFICADO.

4) los resultados parciales de estimación (Ver 3.6.2.3)

5) los intervalos de confianza

6) los resultados de determinación de tamaño de muestra

Para imprimir la información arriba mencionada, únicamente se requiere contestar 'SI' cada vez que aparezcan en pantalla las siguientes preguntas:

?Quiere imprimir los datos? (SI o NO) _	(I-20)
?Quiere imprimir los resultados? (SI o NO) _	(I-21)
?Quiere imprimir los resultados parciales? (SI o NO) _	(I-22)

Nota: Se debe confirmar que la microcomputadora que se esta utilizando esté conectada con la impresora y ésta a su vez esté encendida.

3.6. ORGANIZACION DE LOS MODULOS

Todos los módulos de este paquete están compuestos por tres secciones principales. En la SECCION I, el módulo solicita información general como el total de unidades en la población (N), el tamaño de muestra (n), los valores de la característica de interés (y) para cada elemento de la muestra,etc.. En la SECCION II, se realizan los procedimientos necesarios para obtener intervalos de confianza. En la SECCION III, se calcula el tamaño de muestra para estimar los parámetros bajo diferentes requerimientos como precisión y confianza, varianza mínima etc..

Como ya se mencionó en 3.3, existen tres alternativas a elegir. Cuando se ejige la alternativa [1], en el módulo se ejecuta la SECCION I y la SECCION II. En el caso de la alternativa [2], en el módulo se ejecuta la SECCION I y la SECCION III. En el caso de la alternativa [3] en el módulo se ejecuta la SECCION I, la SECCION II y la SECCION III.

A continuación, se presentan las especificaciones de cada sección.

3.6.1 SECCION I -INFORMACION GENERAL-

En esta sección, se debe introducir la información general que se requiere en cada módulo, es decir, se captura la información que se necesita para cada tipo de diseño de muestra.

1) M.A.S

a) Total de unidades en la población (N)

Si la población que tiene es infinita, se debe introducir 9999999999.

b) Total de la variable auxiliar (X)

Este valor se necesita únicamente cuando se quiere estimar vía razón o regresión o determinar un tamaño de muestra para estimar vía razón o regresión una media o un total.

La información que a continuación solicita el módulo debera introducirse únicamente en el caso en el que ya se disponga de información de una muestra y el problema a resolver sea la estimación de algún parámetro. En consecuencia también deberá introducirse en el caso en el que se disponga de los datos por unidad de una muestra piloto.(Ver 3.3)

c) Tamaño de muestra (n)

Se permite introducir un número máximo de 50 (por la memoria de la microcomputadora) cuando se quiera estimar cualquier parámetro excepto una proporción. En este ultimo caso no existe límite.

d) Valores de la característica de interés (yi) para cada elemento de la muestra

Cuando el parámetro a estimar es una proporción, en lugar de dar el valor de la variable (y) para cada elemento de la muestra (0 o 1), se deberá proporcionar el total de elementos que poseen la característica de interés.

e) Valores de la variable auxiliar (xi) para cada elemento de la muestra.

Estos valores se necesitan únicamente cuando se quiere obtener los siguientes estimadores:

i) Razón
ii) Media vía razón
iii) Total vía Razón
iv) Media vía regresión
v) Total vía regresión

2) ESTRATIFICADO

a) Número de estratos (L)

Se permite introducir un número máximo de 10. (Por la memoria de la microcomputadora)

b) Total de unidades en cada estrato (Nh)

La información que a continuación solicita el módulo deberá introducirse únicamente en el caso en el que ya se disponga de la información de una muestra y el problema a resolver sea la estimación de algún parámetro. En consecuencia también deberá introducirse en el caso en el que se disponga de los datos por unidad de una muestra piloto.(Ver 3.3)

c) Total de la variable auxiliar en cada estrato (Xh)

Estos valores se necesitan únicamente cuando se quiere estimar algún parámetro mediante estimadores de razón o regresión.

d) Tamaño de muestra por estrato (nh)

Se permite introducir un número máximo de 50 para estimar cualquier parámetro excepto una proporción. En este último caso no existe límite.

e) Valores de la característica de interés (yhi) para cada elemento de la muestra.

Cuando el parámetro a estimar es una proporción, en lugar de dar el valor de la variable (y) para cada elemento de la muestra (0 o 1), se deberá proporcionar el total de elementos por estrato que poseen la característica de interés.

f) Valores de la variable auxiliar (xhi) para cada elemento de la muestra

Estos valores se necesitan únicamente cuando se quiere obtener los siguientes estimadores:

i) Razón
ii) Media vía razón
iii) Total vía razón
iv) Media vía regresión
v) Total vía regresión

3) P.P.T

a) Total de unidades en la población (N)

Si la población que tiene es infinita, se debe introducir 9999999999.

b) Total de la variable auxiliar (X)

La información que a continuación solicita el módulo debe introducirse únicamente en el caso en el que ya se disponga de la información de una muestra y el problema a resolver sea la estimación de algún parámetro. En consecuencia también deberá introducirse en el caso en el que se disponga de los datos por unidad de una muestra piloto. (Ver 3.3)

c) Tamaño de muestra (n)

Se permite introducir un número máximo de 50.

d) Valores de la característica de interés (yi) para cada elemento de la muestra

e) Valores de la variable auxiliar (xi) para cada elemento de la muestra

4) BIETAPICO

a) Total de unidades primarias en la población (N)

Si la población que tiene es infinita, se debe introducir 9999999999.

b) Total de unidades primarias en la muestra (n)

Se permite introducir un número máximo de 10.

c) Total de unidades secundarias en la población (M)

d) Total de unidades secundarias en cada unidad primaria en la muestra (Mi)

e) Total de unidades secundarias muestreadas en cada unidad primaria en la muestra (mi)

Se permite introducir un número máximo de 50

f) Valores de la característica de interés (yij) para cada elemento de la muestra

3.6.2. SECCION II -ESTIMACION-

En esta sección, se obtienen intervalos de confianza para los parámetros de interes. Para construirlos, se debe tener en consideración si la aproximación a la normal de la distribución muestral del estimador es adecuada. Si la aproximación a la normal no es adecuada, se proporciona la alternativa de obtener intervalos de confianza utilizando los percentiles que surgen a partir de la desigualdad de Tchebycheff. Además, se debe verificar si el sesgo del estimador es despreciable en caso de que éste sea un estimador sesgado pero consistente.

3.6.2.1. LIMITES DE CONFIANZA

1) M.A.S

a) Cuando se quiere construir un intervalo de confianza para una media y un total mediante estimadores simples, se deberá tomar en cuenta la distribución de la variable (y) y el tamaño de muestra.

Acerca de la distribución de la variable (y), el módulo pregunta:

(I-23)

```
?La caracteristica tiene distribucion normal? (SI o NO)
_
```

Si la variable (y) se distribuye normalmente, se deberá contestar un 'SI' y en caso contrario un 'NO'

Acerca del tamaño de muestra, el programa compara el tamaño de muestra que se ha introducido con el número 30.(Tamaño de muestra mínima que empíricamente se ha considerado adecuado para aproximar la distribución muestral con la distribución normal)

Tomando estas dos condiciones, el programa tomará una decisión acerca de los porcentiles a utilizar para el calculo de los limites de confianza.

i) Si el tamaño de muestra es mayor o igual a 30, se utilizan los percentiles de la distribución normal para calcular el intervalo de confianza

ii) Si el tamaño de muestra es menor de 30, pero la variable (y) se distribuye normalmente, se utilizan los percentiles de la distribución la t de Student para calcular el intervalo de confianza.

iii) Si el tamaño de muestra es menor de 30 y la variable (y) no se distribuye normalmente, se utilizan los percentiles que se originan a partir de la desigualdad de Tchebycheff, para calcular el intervalo de confianza.

b) Cuando se quiere construir un intervalo de confianza para una proporción, se compara la proporción estimada y el tamaño de muestra que se ha introducido, con el siguiente cuadro:(Cochran 1974 "Técnicas de Muestreo" Página 89)

Valores más pequeños de np para uso de la aproximación normal

p	np = Numero observado en la clase mas pequena	n=Tamano de muestra
0.5	15	30
0.4	20	50
0.3	24	80
0.2	40	200
0.1	60	600
0.05	70	1400
~0.	80	∞

Si la proporción estimada es mayor de 0.5, se usará (1-p) para leer la tabla.

i) Si el tamaño de muestra es mayor o igual a la n del cuadro que corresponde a la proporción estimada, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si el tamaño de muestra es menor de la n del cuadro que corresponde a la proporción estimada, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

c) Cuando se quiere construir un intervalo de confianza para una razón o para algún parámetro estimado mediante estimadores de razón o regresión, se toma en cuenta únicamente el tamaño de muestra.

i) Si el tamaño de muestra es mayor o igual a 30, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si el tamaño de muestra es menor de 30, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

2) ESTRATIFICADO

a) Cuando se quiere construir un intervalo de confianza para una media o un total mediante estimadores simples, también se debe tener en consideración la distribución de la variable (y) y el tamaño de muestra haciendo estas consideraciones por estrato.

Acerca de la distribución de la variable (y), el módulo pregunta:

(I-24)

```
?La caracteristica en todos los estratos tiene distribucion normal?
(SI o NO)
_
```

Respecto al tamaño de muestra, se compara el tamaño de muestra de cada estrato con el número 30.

i) Si en todos los estratos el tamaño de muestra es mayor o igual a 30, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si al menos un tamaño de muestra por estrato es menor de 30, pero la variable (y) se distribuye normalmente en todos los estratos, se utiliza la distribución t de Student para calcular el intervalo de confianza, con la modificación en los grados de libertad propuesta por Cochran. (1974 "Técnicas de Muestreo" Página 136)

iii) Con que para un solo estrato ocurra que el tamaño de muestra sea menor de 30 y la variable (y) no se distribuya normalmente, se utilizará la desigualdad de Tchebycheff para calcular el intervalo de confianza.

b) Cuando se quiere construir un intervalo de confianza para una proporción, se compara en todos los estratos la proporción estimada y el tamaño de muestra por estrato con el cuadro de "Valores más pequeños de np para uso de la aproximación normal".

i) Si en todos los estratos el tamaño de muestra es mayor o igual a la n del cuadro que corresponde a la proporción estimada para cada estrato, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si al menos en un estrato el tamaño de muestra es menor de la n del cuadro que corresponde a la proporción estimada para ese estrato, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

c) Cuando se quiere construir un intervalo de confianza para una razón o para algún parámetro estimado mediante estimadores combinados de razón o regresión, se toma en cuenta el tamaño de muestra total.

i) Si el tamaño de muestra total es mayor o igual a 30, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si el tamaño de muestra total es menor de 30, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

d) Cuando se quiere construir un intervalo de confianza para algún parámetro, estimado mediante estimadores separados de razón o regresión, se toma en cuenta el tamaño de muestra por estrato en todos los estratos.

i) Si el tamaño de muestra por estrato es mayor o igual a 30 en todos los estratos, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si al menos un tamaño de muestra por estrato es menor de 30, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

3) P.P.T

En P.P.T se tiene en consideración solamente el tamaño de muestra.

i) Si el tamaño de muestra es mayor o igual a 30, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si el tamaño de muestra es menor de 30, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

4) BIETAPICO

En BIETAPICO se deben tomar en cuenta tres condiciones para decidir el percentil a utilizar en el cálculo de intervalos de confianza.

a) Distribución de la variable (y)

b) Total de unidades primarias en la muestra

c) Total de unidades secundarias muestreadas por unidad primaria en la muestra

i) Si la variable (y) se distribuye normalmente o el total de unidades primarias en la muestra es mayor o igual a 30 o el total de unidades secundarias muestreadas por unidad primaria es mayor o igual a 30 en todas las unidades primarias en la muestra, se utiliza la aproximación normal para calcular el intervalo de confianza.

ii) Si no se cumple ninguna de las condiciones de arriba, se utiliza la desigualdad de Tchebycheff para calcular el intervalo de confianza.

Después de que el programa escoge un percentil, aparece en pantalla uno de los siguientes mensajes dependiendo del percentil escogido:

1) | Siendo estricto, se usa el percentil *Normal* | (I-25)

2) | Siendo estricto, se usa el percentil *t Student* | (I-26)

3) | Siendo estricto, se usa el percentil *Tchebycheff* | (I-27)

Pero en algunas ocasiones, el usuario puede no estar de acuerdo con la decisión del programa. Por lo tanto el programa pregunta acerca del percentil que se ha escogido:

```
?Esta de acuerdo? (SI o NO)
_
```

(I-28)

Si no está de acuerdo, se debe contestar 'NO'.

En el caso de 'NO', se vuelve a preguntar sobre percentil:

```
?Que tipo de percentil quiere usar?
(Z:Normal o T:t Student o K:Tchebycheff)
_
```

(I-29)

Se debe introducir una de las letras 'Z' o 'T' o 'K' dependiendo del percentil que se elija.

3.6.2.2. SESGO DEL ESTIMADOR

Si el parámetro a estimar es una razón o algún parámetro mediante estimadores de razón o regresión, se debe evaluar si el sesgo del estimador es despreciable, ya que los estimadores de estos parámetros son sesgados pero consistentes.

A) Sesgo de estimadores de razón

El sesgo estandarizado de un estimador de razón se encuentra acotado por el coeficiente de variación del estimador de la media de la variable auxiliar. (Ver 2.3)

A continuación se describe cómo decide el programa si el sesgo es despreciable o no.

1) M.A.S

Se calcula el coeficiente de variación muestral de $\hat{\bar{X}}$ y aparece el resultado del cálculo en pantalla como sigue:

(I-30)

```
* Coeficiente de variacion del estimador de la media (x) = resultado del
  cálculo
```

se toma una decisión con base en lo siguiente:

a) Si el resultado del cálculo es menor o igual a 0.1, se considera el sesgo del estimador despreciable.

b) Si el resultado del cálculo es mayor de 0.1, no se considera el sesgo del estimador despreciable.

2) ESTRATIFICADO

i) Cuando se estima una razón o un parámetro mediante estimadores combinados, se estima el coeficiente de variación de $\hat{\bar{X}}$ utilizando desde luego el estimador de la varianza de $\hat{\bar{X}}$ en ESTRATIFICADO. Aparece el resultado del cálculo en pantalla como sigue:

(I-32)

```
* Coeficiente de variacion del estimador de la media (x) = resultado
  del cálculo
```

Se toma una decisión con base en lo siguiente:

a) Si el resultado del cálculo es menor o igual a 0.1, se considera el sesgo del estimador despreciable.

b) Si el resultado del cálculo es mayor de 0.1, no se considera el sesgo del estimador despreciable.

ii) Cuando se estima un parámetro mediante estimadores separados, primero se estima para cada estrato el coeficiente de variación de $\hat{\bar{X}}$ y posteriormente se escoge el mayor de todos los coeficientes de variación estimados y se multiplica por la raíz cuadrada del número de estratos (Ver 2.3). Aparece el resultado del cálculo en pantalla como sigue:

```
* SQRT(L)xCoeficiente de variacion = resultado del cálculo
```
(I-32)

Se toma una decisión con base en lo siguiente:

a) Si el resultado del cálculo es menor o igual a 0.3, se considera el sesgo del estimador despreciable.

b) Si el resultado del cálculo es mayor de 0.3, no se considera el sesgo del estimador despreciable.

B) Sesgo de estimadores de regresión

En el caso de estimadores de regresión, si la relación entre la variable (y) y la variable (x) en la población es exactamente lineal, el estimador de regresión es insesgado, pero en gran número de casos, es muy difícil tener una población así. Por lo tanto el estimador de regresión se considera sesgado pero consistente. Se decide si el sesgo del estimador es despreciable de acuerdo a lo siguiente(Ver 2.3):

1) M.A.S

a) Si el tamaño de muestra es mayor o igual a 50, se considera el sesgo del estimador despreciable.

b) Si el tamaño de muestra es menor de 50, no se considera el sesgo del estimador despreciable.

2) ESTRATIFICADO

i) En el caso de estimadores combinados

a) Si el tamaño de muestra total es mayor o igual a 50, se considera el sesgo del estimador despreciable.

b) Si el tamaño de muestra total es menor de 50, no se considera el sesgo del estimador despreciable.

ii) En el caso de estimadores separados

a) Si el tamaño de muestra por estrato es mayor o igual a 50 en todos los estratos, se considera el sesgo del estimador despreciable.

b) Si al menos un tamaño de muestra por estrato es menor de 50, no se considera el sesgo del estimador despreciable.

Después de tomar una decisión acerca del sesgo del estimador, aparece uno de los siguientes mensajes en pantalla:

1) | Siendo estricto, el sesgo es despreciable | (I-33)

2) | Siendo estricto, el sesgo no es despreciable | (I-34)

En algunas ocasiones, el usuario puede no estar de acuerdo con la decisión del programa, por lo tanto se pregunta:

| ?Esta de acuerdo? (SI o NO)
_ | (I-35)

Si no está de acuerdo, debe contestar 'NO'.

En el caso de 'NO', se vuelve a preguntar:

| ?El sesgo es despreciable? (SI o NO)
_ | (I-36)

Si el usuario opina que el sesgo es despreciable, debe contestar 'SI', en caso contrario, debe contestar 'NO'

Al final de esta parte, si el sesgo del estimador no es despreciable, aparece el siguiente mensaje en pantalla:

¡El sesgo no es despreciable, por lo tanto no se puede calcular el intervalo de confianza!
¡Hay que aumentar el tamano de muestra! (I-37)

Con este mensaje termina el módulo en curso.

3.6.2.3. RESULTADOS

En esta sección se proporcionan no solo intervalos de confianza sino también los resultados parciales que se consideran importantes. A continuación se presentan los resultados parciales que aparecen en pantalla.

1) M.A.S

a) Estimador

b) Estimador de la varianza del estimador o en su caso, el estimador de la aproximación del error cuadrático medio del estimador

c) Varianza muestral de yi: sy^2

Este valor aparece en pantalla cuando se utilizan los siguientes estimadores:

i) Media simple
ii) Total simple
iii) Media vía regresión
iv) Total vía regresión

d) Estimador de Su^2 : su^2

Este valor aparece en pantalla cuando se ha estimado una razón o se han utilizado estimadores de razón.

e) Estimador de una razón

Este valor aparece en pantalla cuando se han utilizado estimadores de razón.

f) Estimador del coeficiente de regresión

Este valor aparece en pantalla cuando se han utilizado estimadores de regresión.

2) ESTRATIFICADO

Dependiendo del parámetro y del estimador elegido para estimarlo se proporciona la siguiente información para cada estrato:

Media Total Estimador simple	Proporcion	Razon	Media y total Estimador de Razon y Regresion
h	h	h	h
N	N	N	N
W	W	W	W
n	n	n	n
My	y	Ty	r
Vy	p	Tx	b
			CV

h : Número de estrato
N = Nh : Total de unidades por estrato
W = Wh : Proporción de la población en el estrato
n = nh : Tamaño de muestra por estrato
My = $\bar{y}_h$: Media muestral de la variable (y) por estrato
Vy = s^2_{yh} : Varianza muestral de la variable (y) por estrato
y = $\sum y_{hi}$: Total de elementos de la muestra que poseen la característica de interés por estrato
p = ph : Proporción muestral por estrato
Ty = $\hat{Y}_h$: Estimador del total de la variable(y) por estrato
Tx = $\hat{X}_h$: Estimador del total de la variable (x) por estrato
r = $\hat{R}_h$: Estimador de la razón por estrato
b = $\hat{b}_h$: Estimador del coeficiente de regresión por estrato
CV = $CV(\hat{X})$: Coeficiente de variación de X muestral por estrato

Además se proporcionan los siguientes resultados:

a) Estimador

b) Estimador de la varianza del estimador o en su caso, el estimador de la aproximación del error cuadrático medio del estimador

c) Estimador de la razón

Este valor aparece cuando se han utilizado estimadores combinados de razón

d) Estimador del coeficiente de regresión

Este valor aparece cuando se han utilizado estimadores combinados de regresión.

3) P.P.T

a) Estimador

b) Estimador de la varianza del estimador

4) BIETAPICO

Se proporciona la siguiente información para cada unidad primaria:

i		: Número de unidad primaria
Ni		: Total de unidades secundarias en la i-ésima unidad primaria en la muestra
mi		: Total de unidades secundarias muestreadas en la i-ésima unidad primaria en la muestra
My	$= \bar{y}i$	: Media muestral de la variable (y) para la i-ésima unidad primaria en la muestra
Ty	$= \hat{Y}i$	: Estimador del total de la variable (y) en la i-ésima unidad primaria en la muestra

Además se proporcionan los siguientes resultados:

a) Estimador

b) Estimador de la varianza del estimador

Al final aparece el intervalo de confianza calculado con el valor de Z, t o K que se haya escogido dependiendo de la aproximación que se elija como correcta. Se pueden construir intervalos de confianza varias veces con diferentes percentiles y diferentes niveles de confianza.

Nota: El valor de Z o t se selecciona de las tablas de la normal estandar o de la t de Student que se encuentran grabados en un archivo en el disco de la UNIDAD-A.

3.6.3. SECCION III -DETERMINACION DEL TAMANO DE MUESTRA-

En esta sección se determina el tamaño de muestra. Se proporcionan las siguientes alternativas para cada tipo de diseño de muestra.

1) M.A.S y P.P.T

a) Determinar el tamaño de muestra que se requiere para tener una precisión y confianza prefijadas.

b) Determinar el tamaño de muestra que se requiere para obtener un coeficiente de variación de estimador prefijado.

2) ESTRATIFICADO

Para este tipo de muestreo se proporcionan tres tipos de asignación para distribuir el tamaño de muestra total a cada estrato y por cada tipo de asignación se presentan dos alternativas para determinar el tamaño de muestra total.

i) Asignación óptima y de Neyman

a) Determinar el tamaño de muestra que minimice la varianza del estimador para un costo total fijo.

b) Determinar el tamaño de muestra que minimice el costo total para una varianza fija del estimador.

ii) Asignación proporcional

a) Determinar el tamaño de muestra que se requiere para tener una precisión y confianza prefijadas.

b) Determinar el tamaño de muestra que se requiere para obtener un coeficiente de variación del estimador prefijado.

Nota: En este paquete no se incluyen los procedimientos para la determinación del tamaño de muestra para un BIETAPICO.

Los procedimientos de esta sección se pueden repetir tantas veces como uno desee respondiendo un 'SI' a la pregunta:

(I-38)

```
?Quiere determinar otro tamano de muestra?
 (SI o NO)
_
```

A continuación, se presentan algunas consideraciones que se deben tomar en cuenta para la determinación del tamaño de muestra.

3.6.3.1. APROXIMACION A LA DISTRIBUCION MUESTRAL DE UN ESTIMADOR

Cuando se quiere determinar el tamaño de muestra para estimar una media, un total o una proporción fijando precisión y confianza, se tiene que decidir con qué distribución se aproximará la distribución muestral del estimador para poder obtener los percentiles, que aseguren para un determinado tamaño de muestra la precisión y confianza deseadas.

a) Cuando se quiere determinar un tamaño de muestra para estimar una media o un total en M.A.S fijando precisión y confianza, se debe investigar si la distribución de la variable (y) es normal o no. Por lo tanto el programa pregunta:

```
?Se puede suponer distribucion normal para la variable (y)?
(SI o NO)
_
```
(I-39)

Si se puede considerar que la variable (y) se distribuye normalmente, se debe contestar 'SI'.

Si se contesta 'SI', aparece el siguiente mensaje en pantalla:

i)
```
Se usa la distribucion normal para calcular el tamano
de muestra.
```
(I-40)

Si se contesta 'NO', aparece el siguiente mensaje en pantalla:

ii)
```
Se usa la desigualdad Tchebycheff para calcular el tamano
de muestra.
```
(I-41)

b) Cuando se quiere determinar el tamaño de muestra para estimar una proporción en M.A.S fijando precisión y confianza, no es bueno utilizar la aproximación normal a la distribución muestral del estimador si se conoce que la proporción de elementos es muy pequeña o muy grande, ya que en este caso la aproximación es mala. Por lo tanto, el programa compara la información que tiene acerca de la proporción con el valor de 0.1.

i)
```
La proporcion no es tan pequena (p>0.1), por lo tanto
se usa la distribucion normal para calcular el tamano
de muestra.
```
(I-42)

ii)
```
La proporcion es demasiado pequena (p<0.1), por lo
tanto, se usa la desigualdad Tchebycheff para calcular
el tamano de muestra.
```
(I-43)

Después de que aparece el mensaje sobre la distribución que se ha escogido, el programa pregunta:

```
?Esta de acuedo? (SI o NO)
_
```
(I-44)

Si no se está de acuerdo con la distribución que el programa ha escogido, se debe contestar 'NO'. Si contestó 'NO', aparece el siguiente pregunta:

```
?Que tipo de distribucion quiere usar?          (I-45)
(Z:Normal o K:Tchebycheff)
_
```

Si se quiere usar la distribución normal, se debe introducir 'Z'. Si se quiere usar la desigualdad Tchebycheff, se debe introducir 'K'.

En el caso de determinación del tamaño de muestra para estimar una media o un total, si se ha escogido la distribución normal como aproximación, entonces si el tamaño de muestra resultante es menor de 30, se realiza un procedimiento iterativo utilizando la distribución de la t de Student el cual termina cuando la parte entera del resultado del cálculo no sea diferente entre un paso y el siguiente. (Ver 2.3)

c) Cuando se quiere determinar el tamaño de muestra para estimar los siguientes parámetros fijando precisión y confianza, se usa la distribución normal para calcular el tamaño de muestra:

1) M.A.S

i) Razón
ii) Media vía razón
iii) Total vía razón
iv) Media vía regresión
v) Total vía regresión

2) ESTRATIFICADO con asignación proporcional

i) Media
ii) Total
iii) Proporción

3) P.P.T

i) Media
ii) Total

3.6.3.2. PARA ESTIMADORES SESGADOS

Los siguientes estimadores son sesgados pero consistentes:

1) Razón
2) Media vía razón
3) Total vía razón
4) Media vía regresión
5) Total vía regresión

Cuando se quiere determinar el tamaño de muestra para los estimadores de razón, se toma en cuenta que el tamaño de muestra asegure que el sesgo del estimador sea despreciable. Por lo tanto primero se calcula un tamaño de muestra mediante el procedimiento que se haya escogido, y segundo se calcula un tamaño de muestra que cumpla con que el coeficiente de variación de $\bar{X}$ sea de 0.1. Posteriormente se escoge el mayor de los dos. En el caso de los estimadores de regresión, en este paquete no se hace consideraciones que garantice que el sesgo del estimador sea despreciable ya que si el tamaño de muestra es grande ($n \geq 50$), se considera que éste es despreciable, lo cual deberá verificar el usuario.

3.6.3.3. INFORMACION QUE SE NECESITA

1) Cuando se quiere determinar el tamaño de muestra fijando precisión y confianza, se deberá introducir:

   a) Precisión deseada
   b) Confianza deseada

2) Cuando se quiere determinar el tamaño de muestra fijando el coeficiente de variación, se debera introducir:

   a) Coeficiente de variación del estimador deseado

3) Cuando se quiere determinar el tamaño de muestra usando la asignación óptima,

   i) Si se quiere minimizar la varianza para un costo total fijo, se deberá introducir:

      a) Costo total
      b) Costo fijo o administrativo
      c) Costo por unidad para todos los estratos

   ii) Si se quiere minimizar el costo total para una varianza fija, se deberá introducir:

      a) Varianza deseada
      b) Costo por unidad para todos los estratos

4) Cuando se quiere determinar el tamaño de muestra usando la asignación de Neyman

   i) Si se quiere minimizar la varianza para un costo total fijo, se deberá introducir:

      a) Costo total
      b) Costo fijo o administrativo
      c) Costo por unidad

   ii) Si se quiere minimizar el costo total para una varianza fija, se deberá introducir:

a) Varianza deseada

Además se necesita la siguiente información si se escogió [2] en I-2 o I-3 o I-5 de 3.2.

1) M.A.S

i) Cuando se quiere determinar el tamaño de muestra para estimar una media o un total mediante estimadores simples, se introduce:

a) Varianza de yi: Sy^2, en el caso en que se desee una precisión y confianza específicas para el estimador

b) Varianza de yi: Sy^2 y la media de yi: $\bar{Y}$, en el caso en que se desee un coeficiente de variación de $\hat{\bar{Y}}$ específico para el estimador

ii) Cuando se quiere determinar un tamaño de muestra para estimar una proporción, el programa solicita información acerca de la proporcion

iii) Cuando se quiere determinar el tamaño de muestra para estimar una razón o algún parámetro mediante estimadores de razón, se deberá proporcionar:

a) Valor de Su^2, en el caso en que se desee una precisión y confianza específicas para el estimador

b) Valor de Su^2 y la media de yi: $\bar{Y}$ o la razón:R, en el caso en que se desee un coeficiente de variación específico para el estimador

c) Varianza de xi: Sx^2 y la media de xi: $\bar{X}$. Esto lo solicita el programa para determinar el tamaño de muestra que garantice un sesgo despreciable para el estimador.

iv) Cuando se quiere determinar el tamaño de muestra para estimar algún parámetro mediante estimadores de regresión, se deberá proporcionar:

a) Varianza de yi: Sy^2 y el coeficiente de correlación entre (x) y (y): ρxy, en el caso en que se desee una precisión y confianza específicas para el estimador

b) Varianza de yi: Sy^2, el coeficiente de correlación entre (x) y (y): ρ xy y la media de yi: $\bar{Y}$, en el caso en que se desee un coeficiente de variación para el estimador

2) ESTRATIFICADO

i) Cuando se quiere determinar el tamaño de muestra para estimar una media o un total, se deberá proporcionar:

a) Varianza de yi por estrato : S^2_{yh} para todas las asignaciones.

b) Media de yi: $\bar{Y}$. Esto lo solicita además el programa en el caso en que se desee un coeficiente de variación específico para el estimador

ii) Cuando se quiere determinar el tamaño de muestra para estimar una proporción, se deberá introducir:

a) Información acerca de la proporción por estrato : Ph para todas las asignaciones

b) Información acerca de la proporción : P. Esto lo solicita además el programa en el caso en que se desee un coeficiente de variación específico para el estimador

3) P.P.T

i) Cuando se quiere determinar el tamaño de muestra fijando precisión y confianza, se deberá introducir:

a) Valor de $\sum_{i=1}^{n} p_i \left(Y_i/p_i - Y \right)^2 \qquad p_i = \frac{X_i}{X}$

ii) Cuando se quiere determinar el tamaño de muestra fijando coeficiente de variación, se deberá proporcionar:

a) Valor de $\sum_{i=1}^{n} p_i \left(Y_i/p_i - Y \right)^2$

b) Media de yi: $\bar{Y}$

3.6.3.4. RESULTADOS

Los resultados que aparecen en pantalla son los siguientes:

1) M.A.S

i) Cuando se ha determinado el tamaño de muestra para estimar una proporción o estimar una media o un total mediante estimadores simples o de regresión.

a) Tamaño de muestra

ii) Cuando se ha determinado el tamaño de muestra para estimar una razón o algún parámetro mediante estimadores de razón

a) Tamaño de muestra calculado mediante el método que se haya escogido

b) Tamaño de muestra que cumpla con que el coeficiente de variación de $\hat{\bar{X}}$ sea de 0.1

c) Tamaño de muestra final

2) ESTRATIFICADO

a) Tamaño de muestra por estrato

b) Tamaño de muestra total

3) P.P.T

a) Tamaño de muestra

4. EJEMPLOS

La información que se debe introducir a la computadora por parte del usuario, se indica con una flecha. Los resultados que se obtiene de determinado procedimiento aparecen escritos con minúscula y la información que se necesita para obtenerlos aparecen escrita con mayúscula.

4.1. EJEMPLO 1

Este ejemplo ilustra como determinar el tamaño de muestra cuando se quiere estimar una media en un muestreo aleatorio simple sin reemplazo y se dispone de información de una muestra piloto.

Problema:

Durante 12 meses (de Julio de 1973 a Junio de 1974) fueron capturados 449 peces de una cierta especie en la isla de Rameswaram(India), los cuales fueron numerados. Se quiere determinar el tamaño de muestra que se requiere para estimar la longitud promedio de los peces, con una precisión de 5 centimetros y una confianza de 95%. Como no se disponía de información acerca de como varía dicha longitud, se tomó una muestra piloto de tamaño 10, obteniéndose las siguientes logitudes:

85, 45, 77, 23, 29, 48, 28, 76, 91, 42

(Este ejemplo está tomado de Trejo B (1985), página 23)

Cómputo:

Sección I

Como se trata de un problema de muestreo aleatorio simple sin reemplazo, se elige [1] en I-1 (pagina 62) y aparece en pantalla lo siguiente:

```
    ******************************************
    ** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO **
    ******************************************
    1-Obtener un estimador
    2-Determinar un tamano de muestra
    3-Ambos
    |Numero que corresponde|
--> 3
```

Nota: Se introduce un 3 porque se dispone de los datos vírgenes de una prueba piloto.

```
?Parametro a estimar?
1-Media
2-Total
3-Proporcion
4-Razon
¡Numero que corresponde!
--> 1
```

```
?Que tipo de estimador?
1-Estimador Simple
2-Estimador de Razon
3-Estimador de Regresion
¡Numero que corresponde!
--> 1
```

```
?Total de unidades en la poblacion (N)?
--> 449
```

```
?Tamano de muestra (n)?  n<=50
--> 10
```

```
¡Valores de interes (y)!
?Tiene los datos (y) en un archivo? (SI o NO)
--> NO
```

```
¡Introduzca los datos (y) como se pide!
y[ 1] = 85
y[ 2] = 45
y[ 3] = 77
y[ 4] = 23
y[ 5] = 29  <--
y[ 6] = 48
y[ 7] = 18
y[ 8] = 76
y[ 9] = 91
y[10] = 42
```

Nota: No se tienen los valores de la característica de interés (y) en un archivo. Por lo tanto se deben de introducir tecleando todos los valores.

```
?Quiere ver los datos? (SI o NO)
--> SI
```

```
*** LOS DATOS ***
y[ 1] = 85.00          y[ 2] = 45.00
y[ 3] = 77.00          y[ 4] = 23.00
y[ 5] = 29.00          y[ 6] = 48.00
y[ 7] = 18.00          y[ 8] = 76.00
y[ 9] = 91.00          y[10] = 42.00
```

Nota: Se introdujo el valor de (y) de la observación 7 erroneamente.

```
?Quiere modificar algun dato? (SI o NO)
--> SI
```

```
?Numero de observacion = 7   y[ 7] = 28   <-
```

Nota: Se debe introducir el valor correcto de la observación 7.

```
?Quiere modificar otro dato? (SI o NO)
--> NO
```

```
?Quiere ver los datos otra vez? (SI o NO)
--> SI
```

```
*** LOS DATOS ***
y[ 1] = 85.00          y[ 2] = 45.00
y[ 3] = 77.00          y[ 4] = 23.00
y[ 5] = 29.00          y[ 6] = 48.00
y[ 7] = 28.00          y[ 8] = 76.00
y[ 9] = 91.00          y[10] = 42.00
```

```
?Quiere guardar los datos en un archivo? (SI o NO)
--> SI
```

Nota: Se permite grabar los valores de la característica de interés en el disco que está colocado en la UNIDAD-B.

```
?Nombre de archivo = b:PECES.DTA <--
```

Nota: Inmediatamente se graban en el archivo que se llama PECES.DTA.

```
?Quiere imprimir los datos? (SI o NO)
--> SI
```

Nota: La impresión de los datos es realizada inmediatamente después de contestar el 'SI'.

Sección II

```
?Quiere imprimir los resultados? (SI o NO)
--> SI
```

Nota: Esta pregunta se refiere a los resultados de estimación.

```
*** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO ***

** MEDIA **

** ESTIMADOR SIMPLE **

* La varianza muestral de (y) = 651.60000

* Estimador = 54.4000

* Estimador de la varianza del estimador = 63.708775
```

Nota: Al mismo tiempo aparecen los resultados en pantalla.

```
    ?La caracteristica tiene distribucion normal? (SI o NO)
--> SI
```

Nota: Se puede considerar que las longitudes de peces se distribuyen normalmente.

```
Siendo estricto (n<30),se usa el percentil *t Student*
```

Nota: Este mensaje aparece en pantalla con base en dos consideraciones: la distribución de la variable (y) y el tamaño de muestra.

```
    ?Esta de acuerdo? (SI o NO)
--> SI
```

Nota: En este caso se considera que la decisión tomada por el programa es la adecuada.

```
    ?Con que confianza (%)?
--> 95
```

```
* El intervalo de confianza al 95% para la media

(        36.3612 ,         72.4388)  t = 2.2600
```

Nota: En este ejercicio el problema a resolver es la determinación del tamaño de muestra y no la estimación de un parámetro. Sin embargo como se disponen de los datos de una prueba piloto se utilizan los procedimientos de estimación para obtener la información que se requiere para determinar el tamaño de muestra. Es únicamente con el fin de ilustrar el uso del paquete que se contesta un SI en esta parte.

```
    ?Quiere probar otro intervalo? (SI o NO)
--> SI
```

Nota: Se permite probar con diferentes percentiles y diferentes niveles de confianza.

```
?Quiere usar otro tipo de percentil? (SI o NO)
--> SI

?Que tipo de percentil quiere usar?
 (Z:Normal o T:t Student o K:Tchebycheff)
--> Z

?Con que confianza (%)?
--> 95

* El intervalo de confianza al 95% para la media

(       38.7557 ,       70.0443)  Z = 1.9600

?Quiere probar otro intervalo? (SI o NO)
--> NO
```

Nota: Terminan los procedimientos de la estimación.

Sección III

```
*** DETERMINACION DEL TAMANO DE MUESTRA ***
1-Fijar precision y confianza
2-Fijar coeficiente de variacion
|Numero que corresponde|
--> 1

?Precision deseada (BO)?
--> 5

?Confianza deseada (%)?
--> 95

?Se puede suponer distribucion normal para la variable (y)
(SI o NO)
--> SI

Se usa la distribucion normal para calcular el tamano de muestra

?Esta de acuerdo? (SI o NO)
--> SI

*** DETERMINACION DEL TAMANO DE MUESTRA ***

** ESTIMACION DE LA MEDIA **

** PRECISION = 5.00 **
** CONFIANZA = 95% **

** VARIANZA DE (Y) = 651.60 **

* Tamano de muestra = 81.87
```

```
    ?Quiere imprimir los resultados? (SI o NO)
--> SI
```

```
    ?Quiere determinar otro tamano de muestra? (SI o NO)
--> NO
```

Nota: Se pueden repetir los procedimientos de la determinación del tamaño de muestra para el mismo problema si se contesta un 'SI', en cuyo caso se empieza de nuevo la Sección III.

```
    ************ Termino *************
    ?Quiere empezar de nuevo? (SI o NO)
--> NO
```

Nota: Se puede utilizar el mismo módulo (M.A.S) con otro problema si se contesta un 'SI', en cuyo caso se inicia nuevamente en la Sección I.

```
**************************
*** Terminamos |GRACIAS| ***
**************************
```

Nota: Termina el módulo.

4.2. EJEMPLO 2

Este ejemplo ilustra cómo estimar un total mediante un estimador simple, un estimador de razón y un estimador de regresión en un muestreo aleatorio simple sin reemplazo y se comparan los tres tipos de estimadores.

Problema: (Este ejemplo fue elabrado por M.V.Z. Jorge Lecumberri)

Después de las fiestas decembrinas, la SSA desea conocer el número total de alcohólicos en la república, para lo cual investigó en cuarenta municipios escogidos aleatoriamente, la cantidad de alcohólicos existentes. El total de municipios es de 2,637 y la población total del país es de 76,791,800. La población total y el número de alcohólicos de cada uno de los municipios muestreados son los siguientes:

Pob.	Alc	Pob.	Alc	Pob.	Alc	Pob.	Alc
125,324	433	209,795	702	189,065	657	324,876	1117
475,689	1512	156,876	572	78,324	297	413,543	1567
89,543	325	413,654	1412	113,234	374	345,638	1267
203,476	789	324,598	1112	225,454	762	143,567	435
108,324	398	654,765	2689	323,657	1294	99,324	303
543,642	1919	145,723	507	524,332	1807	101,345	291
134,809	476	78,098	287	33,454	142	874,543	3309
89,007	312	102,435	338	287,907	909	397,657	1379
54,996	202	354,878	1245	200,657	587	167,843	620
195,876	723	132,209	432	345,002	1269	295,334	920

Cómputo:

** Primera ejecución:estimador simple **

Sección I

Como se trata de un problema de muestreo aleatorio simple sin reemplazo, se elige [1] en I-1 (página 62) y aparece en pantalla lo siguiente:

```
****************************************
** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO **
****************************************
1-Obtener un estimador
2-Determinar un tamano de muestra
3-Ambos
|Numero que corresponde|
--> 1
```

```
?Parametro a estimar?
1-Media
2-Total
3-Proporcion
4-Razon
|Numero que corresponde|
--> 2
```

```
?Que tipo de estimador?
1-Estimador Simple
2-Estimador de Razon
3-Estimador de Regresion
|Numero que corresponde|
--> 1
```

Nota: Primero se estima utilizando solo la información de la variable (y).

```
?Total de unidades en la poblacion (N)?
--> 2637
```

```
?Tamano de muestra (n)?  n<=50
--> 40
```

```
|Valores de interes (y)|
?Tiene los datos (y) en un archivo? (SI o NO)
--> SI
```

```
?Nombre de archivo = b:sregtot.dta <---
```

Nota: Este archivo contiene los valores del número de alcohólicos y además los valores de la población total de los municipios. Pero en este caso, se leen únicamente los valores del número de alcohólicos.

```
?Quiere ver los datos? (SI o NO)
--> SI
```

```
*** LOS DATOS ***
y[ 1] = 433.00          y[ 2] = 1512.00
y[ 3] = 325.00          y[ 4] = 789.00
y[ 5] = 398.00          y[ 6] = 1919.00
y[ 7] = 476.00          y[ 8] = 312.00
y[ 9] = 202.00          y[10] = 723.00
y[11] = 702.00          y[12] = 572.00
y[13] = 1412.00         y[14] = 1112.00
y[15] = 2689.00         y[16] = 507.00
y[17] = 287.00          y[18] = 338.00
y[19] = 1245.00         y[20] = 432.00
y[21] = 657.00          y[22] = 297.00
y[23] = 374.00          y[24] = 762.00
y[25] = 1294.00         y[26] = 1807.00
y[27] = 142.00          y[28] = 909.00
y[29] = 587.00          y[30] = 1269.00
y[31] = 1117.00         y[32] = 1567.00
y[33] = 1267.00         y[34] = 435.00
y[35] = 303.00          y[36] = 291.00
y[37] = 3309.00         y[38] = 1379.00
y[39] = 620.00          y[40] = 920.00
```

```
?Quiere modificar algun dato? (SI o NO)
--> NO
```

```
?Quiere guardar los datos en un archivo? (SI o NO)
--> NO
```

```
?Quiere imprimir los datos? (SI o NO)
--> NO
```

Sección II

```
?Quiere imprimir los resultados? (SI o NO)
--> NO
```

Nota: Esta pregunta se refiere a los resultados de estimación.

```
*** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO ***

** TOTAL **

** ESTIMADOR SIMPLE **

* La varianza muestral de (y) = 469047.691670

* Estimador = 2352929.1750

* Estimador de la varianza del estimador = 80304353683.00000
```

```
?La caracteristica tiene distribucion normal? (SI o NO)
--> no
```

```
Siendo estricto, se usa el percentil *Normal*
```

Nota: Esta decisión se toma con base en dos consideraciones: la distribución de la variable (y) y el tamaño de muestra.

```
?Esta de acuerdo? (SI o NO)
--> SI
```

Nota: En este caso se considera que la decisión tomada por el programa es la adecuada.

```
?Con que confianza (%)?
--> 99
```

```
* El intervalo de confianza al 99% para el total

(  1623225.0876 ,  3082633.2624)  Z = 2.5750
```

```
?Quiere probar otro intervalo? (SI o NO)
--> NO
```

```
************* Termino *************
?Quiere empezar de nuevo? (SI o NO)
--> SI
```

Nota: Se contesta un 'SI', porque se quieren obtener otros estimadores para el parámetro de interés del mismo problema.

** Segunda ejecución:estimador de razón **

Sección I

```
*****************************************
** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO **
*****************************************
1-Obtener un estimador
2-Determinar un tamano de muestra
3-Ambos
!Numero que corresponde!
--> 1
```

```
?Parametro a estimar?
1-Media
2-Total
3-Proporcion
4-Razon
!Numero que corresponde!
--> 2
```

```
?Que tipo de estimador?
1-Estimador Simple
2-Estimador de Razon
3-Estimador de Regresion
|Numero que corresponde|
--> 2
```

Nota: En este caso, se utiliza la información auxiliar (población de municipios) y se estima el total mediante un estimador de razón.

```
?Total de unidades en la poblacion (N)?
--> 2637
```

```
?Total de variable auxiliar (X)?
--> 76791800
```

```
?Tamano de muestra (n)? n<=50
--> 40
```

```
|Valores de interes (y) y Valores auxiliares (x)|
?Tiene los datos (y,x) en un archivo? (SI o NO)
--> SI
```

```
?Nombre de archivo = b:sregtot.dta <--
```

Nota: En este caso se leen los valores del número de alcohólicos y de la población total de los municipios.

```
?Quiere ver los datos? (SI o NO)
--> SI
```

```
*** LOS DATOS ***
y[ 1] =  433.00  x[ 1] = 125324.00 y[ 2] = 1512.00 x[ 2] = 475689.00
y[ 3] =  325.00  x[ 3] =  89543.00 y[ 4] =  789.00 x[ 4] = 203476.00
y[ 5] =  398.00  x[ 5] = 108324.00 y[ 6] = 1919.00 x[ 6] = 543642.00
y[ 7] =  476.00  x[ 7] = 134809.00 y[ 8] =  312.00 x[ 8] =  89007.00
y[ 9] =  202.00  x[ 9] =  54996.00 y[10] =  723.00 x[10] = 195876.00
y[11] =  702.00  x[11] = 209795.00 y[12] =  572.00 x[12] = 156876.00
y[13] = 1412.00  x[13] = 413654.00 y[14] = 1112.00 x[14] = 324598.00
y[15] = 2689.00  x[15] = 654765.00 y[16] =  507.00 x[16] = 145723.00
y[17] =  287.00  x[17] =  78098.00 y[18] =  338.00 x[18] = 102435.00
y[19] = 1245.00  x[19] = 354878.00 y[20] =  432.00 x[20] = 132209.00
y[21] =  657.00  x[21] = 189065.00 y[22] =  297.00 x[22] =  78324.00
y[23] =  374.00  x[23] = 113234.00 y[24] =  762.00 x[24] = 225454.00
y[25] = 1294.00  x[25] = 323657.00 y[26] = 1807.00 x[26] = 524332.00
y[27] =  142.00  x[27] =  33454.00 y[28] =  909.00 x[28] = 287907.00
y[29] =  587.00  x[29] = 200657.00 y[30] = 1269.00 x[30] = 345002.00
y[31] = 1117.00  x[31] = 324876.00 y[32] = 1567.00 x[32] = 413543.00
y[33] = 1267.00  x[33] = 345638.00 y[34] =  435.00 x[34] = 143567.00
y[35] =  303.00  x[35] =  99324.00 y[36] =  291.00 x[36] = 101345.00
y[37] = 3309.00  x[37] = 874543.00 y[38] = 1379.00 x[38] = 397657.00
y[39] =  620.00  x[39] = 167843.00 y[40] =  920.00 x[40] = 295334.00
```

```
?Quiere modificar algun dato? (SI o NO)
--> NO

?Quiere guardar los datos en un archivo? (SI o NO)
--> NO

?Quiere imprimir los datos? (SI o NO)
--> NO
```

Sección II

```
?Quiere imprimir los resultados? (SI o NO)
--> NO

*** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO ***

** TOTAL **

** ESTIMADOR DE RAZON **

* Valor de su2 = 8376.180950

* Estimador de la razon = 0.00354

* Estimador = 271943.5904

* Estimador de la aproximacion del error cuadratico medio =
  1434062696.600000

* Coeficiente de variacion del estimador de la media (x) = 0.1131

Siendo estricto (CV>0.1), el sesgo no es despreciable

?Esta de acuerdo? (SI o NO)
--> NO
```

Nota: El coeficiente de variación de la variable auxiliar es 0.1131, no se considera que esté muy lejano de 0.1. Por lo tanto se puede no estar de acuerdo con la decisión del programa.

```
?El sesgo es despreciable? (SI o NO)
--> SI

Siendo estricto, se usa el percentil *Normal*

?Esta de acuerdo? (SI o NO)
--> SI

?Con que confianza (%)?
--> 99
```

```
* El intervalo de confianza al 99% para el total

(  174430.8632 ,  369456.3177)   Z = 2.5750
```

```
    ?Quiere probar otro intervalo? (SI o NO)
--> NO
```

```
    *********** Termino ***************
    ?Quiere empezar de nuevo? (SI o NO)
--> SI
```

Nota: Se empieza de nuevo para obtener el estimador de regresión.

** Tercera ejecución:estimador de regresión **

Sección I

```
    *****************************************
    ** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO **
    *****************************************
    1-Obtener un estimador
    2-Determinar un tamano de muestra
    3-Ambos
    !Numero que corresponde!
--> 1
```

```
    ?Parametro a estimar?
    1-Media
    2-Total
    3-Proporcion
    4-Razon
    !Numero que corresponde!
--> 2
```

```
    ?Que tipo de estimador?
    1-Estimador Simple
    2-Estimador de Razon
    3-Estimador de Regresion
    !Numero que corresponde!
--> 3
```

Nota: En este caso, se utiliza la información auxiliar y se estima el total mediante un estimador de regresión.

```
    ?Total de unidades en la poblacion (N)?
--> 2637
```

```
    ?Total de variable auxiliar (X)?
--> 76791800
```

```
    ?Tamano de muestra (n)? n<=50
--> 40
```

```
|Valores de interes (y) y Valores auxiliares (x)|
?Tiene los datos (y,x) en un archivo? (SI o NO)
--> SI
```

```
?Nombre de archivo = b:sregtot.dta <--
```

```
?Quiere ver los datos? (SI o NO)
--> NO
```

```
?Quiere modificar algun dato? (SI o NO)
--> NO
```

```
?Quiere guardar los datos en un archivo? (SI o NO)
--> NO
```

```
?Quiere imprimir los datos? (SI o NO)
--> NO
```

Sección II

```
?Quiere imprimir los resultados? (SI o NO)
--> NO
```

```
*** MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO ***

** TOTAL **

** ESTIMADOR DE REGRESION **

* Varianza muestral de (y) = 469047.691670

* Estimador del coeficiente de correlacion = 0.99249

* Estimador del coeficiente de regresion = 0.00374

* Estimador = 152852.2076

* Estimador de la varianza del estimador = 1233923460.30000
```

```
Siendo estricto (n<50), el sesgo no es despreciable
```

Nota: Este mensaje aparece en pantalla porque el tamaño de muestra es 40 y el criterio que se utiliza en este programa es que si el tamaño de muestra es mayor o igual a 50, se considera el sesgo despreciable.

```
?Esta de acuerdo? (SI o NO)
--> NO
```

Nota: No se considera que el tamaño de muestra de este problema sea muy pequeño, por lo tanto no se está de acuerdo con la decisión del programa.

```
     ?El sesgo es despreciable? (SI o NO)
-->  SI

     Siendo estricto, se usa el percentil *Normal*

     ?Esta de acuerdo? (SI o NO)
-->  SI

     ?Con que confianza (%)?
-->  99

     * El intervalo de confianza al 99% para el total

     (  62399.5473 ,  243304.8678) Z = 2.5750

     ?Quiere probar otro intervalo? (SI o NO)
-->  NO

     ********** Termino **************
     ?Quiere empezar de nuevo? (SI o NO)
-->  NO

     *************************
     *** Terminamos |GRACIAS| ***
     *************************
```

Nota: Analizando este problema, se observa que la muestra seleccionada resulta ser una mala muestra, en el sentido de que quedaron seleccionados solamente municipios grandes y entonces donde hay más población puede decirse en general que habrá más alcohólicos. Por lo tanto si se utiliza un estimador simple, se sobreestima el total de alcohólicos. En cambio, en el caso en el que se utiliza la información de la variable (x) para estimar el total, (estimador de razón o de regresión) se corrige al estimador aprovechando la relación entre el número de alcohólicos y la población. Por lo tanto en este caso dado que se tiene un tamaño de muestra relativamente grande, es más adecuado utilizar estimador de razón o de regresión que uno simple.

4.3. EJEMPLO 3

Este ejemplo ilustra como utilizar el paquete para estimar un parámetro en muestreo estratificado aleatorio y como determinar y asignar el tamaño de muestra.

Problema: (Este ejemplo fue elaborado por M.V.Z Jorge Lecumberri)

Una cadena de tiendas de autoservicio desea conocer el total de sus ventas en toda la ciudad en un dia determinado. La cadena tiene 127 tiendas divididas en tres zonas según la clase socioeconómica predominante en el lugar donde están ubicadas: burgueses: 23, clase media: 55, proletariado; 49. Se hizo un muestreo estratificado aleatorio midiendo el total de ventas en cada una de las tiendas seleccionadas. Se obtuvieron los siguientes resultados en millones de pesos:(Dentro de cada estrato se hizo un muestreo aleatorio simple sin reemplazo)

Burgueses	Clase media		Proletariado	
85.4	74.2	75.2	72.2	51.1
94.2	72.7	79.4	64.1	52.7
103.7	89.2	72.1	50.7	57.2
74.2	109.7	92.1	47.2	57.4
93.2	97.2	52.1	84.1	91.2
	74.2	72.4	92.1	59.2
	67.7	67.8	44.1	63.6
	84.2	87.2		

Posteriormente se tomaron los resultados anteriores como un estudio piloto, con finalidad de determinar el tamaño de muestra requerido para minimizar la varianza del estimador dado que se dispone de un presupuesto de 400,000 pesos para el estudio. Los costos para obtener la información son diferentes en cada estrato, ya que las tiendas para la clase burguesa se encuentran concentradas en dos colonias, en cambio para la clase proletaria se encuentran dispersas en la periferia de la ciudad. Por lo que el costo por tienda en la zona burguesa es de 7,500 pesos; para la zona media es de 9,750 pesos y el de la zona proletaria es de 12,300 pesos.

Cómputo:

Sección I

Como se trata de un problema de muestreo estratificado aleatorio, se elige [2] en I-1 (página 62) y aparece en pantalla lo siguiente:

```
*************************************
*** MUESTREO ESTRATIFICADO ALEATORIO ***
*************************************
1-Obtener un estimador
2-Determinar un tamano de muestra
3-Ambos
!Numero que corresponde!
--> 3
```

```
?Parametro a estimar?
1-Media
2-Total
3-Proporcion
!Numero que corresponde!
--> 2
```

```
?Numero de estratos (L)? L<=10
--> 3
```

```
?Total de unidades en cada estrato (Nh)?
Estrato 1 = 23
Estrato 2 = 55 <--
Estrato 3 = 49
```

```
Total de unidades en la poblacion (N) = 127
```

Nota: Este mensaje aparece en pantalla después de introducir todos los totales de unidades en cada estrato.

```
?Quiere ver los datos? (SI o NO)
--> SI
```

```
** TOTAL DE UNIDADES EN CADA ESTRATO **
Estrato 1 = 23
Estrato 2 = 55
Estrato 3 = 49
Total de unidades en la poblacion (N) = 127
```

```
?Quiere modificar algun dato? (SI o NO)
--> NO
```

```
?Tamano de muestra por estrato (nh)? nh<=50
Estrato 1 = 5
Estrato 2 = 16 <--
Estrato 3 = 13
```

```
Tamano de muestra total = 34
```

Nota: Este mensaje aparece en pantalla después de introducir todos los tamaños de muestra en cada estrato.

```
?Quiere ver los datos? (SI o NO)
--> NO
```

```
?Quiere modificar algun dato? (SI o NO)
--> SI
```

Nota: Se introdujo el tamaño de muestra en el estrato 3 erroneamente.

```
?Numero de estrato = 3   Dato = 14 <--
```

```
?Quiere modificar otro dato? (SI o NO)
--> NO
```

```
Tamano de muestra total = 35
```

Nota: Este mensaje aparece en pantalla después de la modificación.

```
?Quiere ver los dato otra vez? (SI o NO)
--> NO
```

```
!Valores de interes (y)!
?Tiene los datos (y) en un archivo? (SI o NO)
```
```
--> SI
```

Nota: Los valores ya estan grabados en el disco que está colocado en la UNIDAD-B.

```
?Nombre de archivo = b:TOTALEST.DTA <--
```

Nota: Se lee el archivo que se llama TOTALEST.DTA.

```
?Quiere ver los datos? (SI o NO)
--> SI
```

```
*** LOS DATOS ***
** Estrato 1 **
y[ 1] =  85.40 [ 2] =  94.20 [ 3] = 103.70 [ 4] =  74.2 [ 5] =  93.2
!Introduzca 'RETURN' _
```

Nota: Cada vez que aparezcan todos los valores de un estrato, se debe teclear 'RETURN'.

```
** Estrato 2 **
y[ 1] =  74.20 [ 2] =  75.20 [ 3] =  72.70 [ 4] =  79.40 [ 5] =  89.20
y[ 6] =  87.20 [ 7] = 109.70 [ 8] =  72.10 [ 9] =  97.20 [10] =  92.10
y[11] =  74.20 [12] =  52.10 [13] =  67.70 [14] =  72.40 [15] =  84.20
y[16] =  67.8
!Introduzca 'RETURN' _
```

```
** Estrato 3 **
y[ 1] =  72.20 [ 2] =  51.10 [ 3] =  64.10 [ 4] =  52.70 [ 5] =  59.70
y[ 6] =  57.20 [ 7] =  47.20 [ 8] =  57.40 [ 9] =  84.10 [10] =  91.20
y[11] =  92.10 [12] =  59.20 [13] =  44.10 [14] =  63.60
!Introduzca 'RETURN' _
```

```
?Quiere modificar algun dato? (SI o NO)
--> NO
```

```
?Quiere guardar los datos en un archivo? (SI o NO)
--> NO
```

```
?Quiere imprimir los datos? (SI o NO)
--> NO
```

Sección II

```
** RESULTADOS DE CADA ESTRATO **
--------------------------------------
h  *          1*          2*          3*
--------------------------------------
N  *         23*         55*         49*
--------------------------------------
W  *  0.181102*   0.433071*   0.385827*
--------------------------------------
n  *          5*         16*         14*
--------------------------------------
My *      90.14*      79.21*      63.99*
--------------------------------------
Vy *     121.57*     187.41*     238.81*
--------------------------------------
|Introduzca 'RETURN'| _
```

Nota: Aparecen los resultados parciales de cada estrato en pantalla y después de revisar, se debe teclear 'RETURN'.

```
    ?Quiere imprimir los resultados parciales? (SI o NO)
--> SI
```

```
*** MUESTREO ESTRATIFICADO ALEATORIO ***

*** TOTAL ***

*** ESTIMADOR SIMPLE ***

* Estimador = 9565.5575

* Estimador de la varianza del estimador = 64445.4936
```

```
    ?Quiere imprimir los resultados? (SI o NO)
--> SI
```

```
    ?La caracteristica de todos los estratos tiene distribucion
    normal? (SI o NO)
--> NO
```

Nota: No se puede considerar que la característica en estudio de este problema tenga distribución normal.

```
Siendo estricto, se usa el percentil *Tchbycheff*
```

Nota: Esta decisión se toma con base en dos condiciones: la distribución de la variable (y) y el tamaño de muestra por estrato en todos los estratos.

```
    ?Esta de acuerdo? (SI o NO)
--> SI
```

```
    ?Con que confianza (%)?
--> 95
```

```
* El intervalo de confianza al 95% para el total

(      8430.2558 ,        10700.8592)  K = 4.4721
```

```
    ?Quiere probar otro intervalo? (SI o NO)
--> NO
```

Seccion III

```
    ** DETERMINACION DEL TAMANO DE MUESTRA **
    1-Asignacion Optima
    2-Asignacion de Neyman
    3-Asignacion Proporcional
    |Numero que corresponde|
--> 1
```

Nota: En este problema se utiliza una distribución óptima, en virtud de que se quiere minimizar la varianza del estimador y los costos varían entre estratos.

```
    1-Minimizar la varianza para un costo total fijo
    2-Minimizar el costo total para una varianza fija
    |Numero que corresponde|
--> 1
```

```
?El costo total = 400000 <--
```

```
?El costo adiministrativo = 0 <--
```

Nota: En este problema no se estipula ningún costo administrativo por lo que se le considero 0.

```
?Los costos por unidad en cada estrato?
Estrato 1 = 7500
Estrato 2 = 9750     <--
Estrato 3 = 12300
```

```
    ?Quiere ver los datos? (SI o NO)
--> SI
```

```
*** LOS COSTOS POR UNIDAD EN CADA ESTRATO ***
Estrato 1 = 7500
Estrato 2 = 9750
Estrato 3 = 12300
```

```
    ?Quiere modificar algun dato? (SI o NO)
--> NO

    ?Quiere imprimir los resultados? (SI o NO)
--> NO

*** DETERMINACION DEL TAMANO DE MUESTRA ***

** ESTIMACION DEL TOTAL **

** LAS VARIANZAS DE (Y) EN CADA ESTRATO **
** ESTRATO 1 = 121.57 **
** ESTRATO 2 = 187.41 **
** ESTRATO 3 = 238.81 **

** ASIGNACION OPTIMA **
** MINIMIZAR LA VARIANZA PARA UN COSTO TOTAL FIJO = 400000 **
** COSTO ADMINISTRATIVO = 0 **
** COSTO POR UNIDAD **
** ESTRATO 1 = 7500 **
** ESTRATO 2 = 9750 **
** ESTRATO 3 = 12300 **

* Tamano de muestra = 38.56
*          Estrato 1 =  6.57
*          Estrato 2 = 17.11
*          Estrato 3 = 15.32

    ?Quiere determinar otro tamano de muestra? (SI o NO)
--> NO

    ************Termino**************
    ?Quiere empezar de nuevo? (SI o NO)
--> NO

    ***************************
    *** Terminamos !GRACIAS! ***
    ***************************
```

4.4. EJEMPLO 4

Este ejemplo ilustra cómo estimar un total mediante estimadores combinados de razón en un muestreo estratificado aleatorio.

Problema: (Este ejemplo fue elaborado por M.V.Z Jorge Lecumberri)

Una empresa que se dedica a hacer quesos, compra leche a establos de diferentes regiones del país y desea conocer el total de grasa contenida en sus compras de leche en un mes. Se sabe que el contenido de grasa en la leche varía según la región donde se compre la misma, por lo cual se decidió hacer un muestreo estratificado aleatorio de establos. Las compras de leche se hacen a establos de siete regiones diferentes. Para cada establo en muestra se midió el total de litros de leche que se compraron así como el contenido de grasa en ellos:

Región 1

Lt.	Grasa
19310	679
21415	750
18325	640
22122	737
19103	670

Región 2

Lt.	Grasa
27225	790
28322	829
21427	615
22196	651
25722	751
27124	764
25684	745

Región 3

Lt.	Grasa
21217	679
22727	727
24545	790
25122	807
23234	744
24199	767
21122	669

Región 4

Lt.	Grasa
21912	591
20314	547
22189	602
24702	670
23145	623
22203	597
24924	670
23432	629
23392	630
21812	588
19407	524
27322	738
24212	652
16744	451
23194	624
22556	608
21921	589
19894	537

Región 5

Lt.	Grasa
18924	624
17314	569
21912	725
24515	809
22124	734
21914	721
22115	719
19814	655

Región 6

Lt.	Grasa
21917	791
22812	820
18214	650
19715	708

Región 7

Lt.	Grasa
19124	637
19722	652
21912	725
22814	749
23715	784
24812	819
18917	621
20812	686

El número de establos a los que se les compra leche por región y el número de litros comprados se dan a continuación:

Región	Establos	Litros
1	25	500000
2	34	850000
3	37	851000
4	105	2310000
5	44	836000
6	22	440000
7	42	882000

Cómputo:

Sección I

Como se trata de un problema de muestreo estratificado aleatorio, se elige [2] en I-1 (pagina 62) y aparece en pantalla lo siguiente:

```
***********************************
*** MUESTREO ESTRATIFICADO ALEATORIO ***
***********************************
1-Obtener un estimador
2-Determinar un tamano de muestra
3-Ambos
|Numero que corresponde|
--> 1
```

```
?Parametro a estimar?
1-Media
2-Total
3-Proporcion
4-Razon
|Numero que corresponde|
--> 2
```

```
?Que tipo de estimador?
1-Estimador Simple
2-Estimador de Razon
3-Estimador de Regresion
|Numero que corresponde|
--> 2
```

Nota: Se elige un estimador de razón porque se dispone de información auxiliar que se supone se encuentra relacionada aproximadamente en forma proporcional con la variable de interés.

```
?Numero de estratos (L)? L<=10
--> 7
```

```
?Total de unidades en cada estrato (Nh)?
Estrato  1 = 25
Estrato  2 = 34
Estrato  3 = 37
Estrato  4 = 105   <--
Estrato  5 = 44
Estrato  6 = 22
Estrato  7 = 42
```

```
Total de unidades en la poblacion (N) = 309
```

```
?Quiere ver los datos? (SI o NO)
--> NO
```

```
    ?Quiere modificar algun dato? (SI o NO)
--> NO

    ?Total de variable auxiliar en cada estrato (Xh)?
    Estrato  1 = 500000
    Estrato  2 = 850000
    Estrato  3 = 851000
    Estrato  4 = 2310000   <--
    Estrato  5 = 836000
    Estrato  6 = 440000
    Estrato  7 = 882000

    Total de la variable auxiliar en la poblacion (X) = 6669000.00

    ?Quiere ver los datos? (SI o NO)
--> NO

    ?Quiere modificar algun dato? (SI o NO)
--> NO

    ?Tamano de muestra por estrato (nh)? nh<=50
    Estrato  1 = 5
    Estrato  2 = 7
    Estrato  3 = 7
    Estrato  4 = 18   <--
    Estrato  5 = 8
    Estrato  6 = 4
    Estrato  7 = 8

    Tamano de muestra total (n) = 57

    ?Quiere ver los datos? (SI o NO)
--> NO

    ?Quiere modificar algun dato? (SI o NO)
--> NO

    |Valores de interes (y) y Valores auxiliares (x)|
    ?Tiene los datos (x,y) en un archivo? (SI o NO)
--> SI

    ?Nombre de archivo = b:RZTOCO.DTA <--

    ?Quiere ver los datos? (SI o NO)
--> NO

    ?Quiere modificar algun dato? (SI o NO)
--> NO

    ?Quiere guardar los datos en un archivo? (SI o NO)
--> NO

    ?Quiere imprimir los datos? (SI o NO)
--> NO
```

Sección II

```
** RESULTADOS PARCIALES DE CADA ESTRATO **
-----------------------------------------------
h  *        1*        2*        3*        4*        5*
-----------------------------------------------
N  *       25*       34*       37*      105*       44*
-----------------------------------------------
W  *  0.0809*  0.1100*  0.1197*  0.3398*  0.1424*
-----------------------------------------------
n  *        5*        7*        7*       18*        8*
-----------------------------------------------
r  *   0.035*   0.029*   0.032*   0.027*   0.033*
-----------------------------------------------
b  *   0.028*   0.029*   0.033*   0.027*   0.033*
-----------------------------------------------
CV *    0.03*    0.03*    0.02*    0.02*    0.03*
-----------------------------------------------
X  *  500000*  850000*  851000* 2310000*  836000*
-----------------------------------------------
|Introduzca 'RETURN'| _
```

Nota: Como toda la información no puede desplegarse en la pantalla (se organizó de tal forma que aparezcan a lo más 5 estratos), se introduce un 'RETURN' y aparece la información restante.

```
----------------------
h  *        6*        7*
----------------------
N  *       22*       42*
----------------------
W  *  0.0712*  0.1359*
----------------------
n  *        4*        8*
----------------------
r  *   0.036*   0.033*
----------------------
b  *   0.037*   0.033*
----------------------
CV *    0.04*    0.03*
----------------------
X  *  440000*  882000*
----------------------
|Introduzca 'RETURN'| _
```

```
?Quiere imprimir los resultados parciales de cada estrato?
(SI o NO)
--> NO
```

```
En el caso de estimadores de razon y regresion
1-Estimador Separado
2-Estimador Combinado
|Numero que corresponde|
--> 2
```

Nota: Se elige un estimador combinado ya que el número de estratos es relativamente grande, los tamaños de muestra son pequeños y las razones por estrato son parecidas.

```
*** MUESTREO ESTRATIFICADO ALEATORIO ***

** TOTAL **

** ESTIMADOR DE RAZON **

** ESTIMADOR COMBINADO **

* Estimador de la razon = 0.03060

* Estimador = 204053.4237

* Estimador de la aproximacion del error cuadratico medio = 113836.5069
```

```
?Quiere imprimir los resultado? (SI o NO)
--> NO
```

```
* Coeficiente de variacion del estimador de la media de (x) = 0.0119
```

```
Siendo estricto (CV<=0.1), el sesgo es despreciable
```

```
?Esta de acuerdo? (SI o NO)
--> SI
```

```
Siendo estricto, se usa el percentil *Normal*
```

Nota: Esta decisión se toma con base en el tamaño de muestra total.

```
?Esta de acuerdo? (SI o NO)
--> SI
```

```
?Con que confianza? (%)
--> 95
```

```
* El intervalo de confianza al 95% para el total

(     203392.1262 ,      204714.7211)  Z = 1.9600
```

```
?Quiere probar otro intervalo? (SI o NO)
--> NO
```

```
************Termino**************
?Quiere empezar de nuevo? (SI o NO)
--> NO
```

```
***************************
*** Terminamos !GRACIAS! ***
***************************
```

4.5. EJEMPLO 5

Este ejemplo ilustra cómo estimar un total en un muestreo con probabilidad proporcional al tamaño con reemplazo.

Problema: (Este ejemplo fue elaborado por M.V.Z. Jorge Lecumberri)

Se desea conocer el total de operaciones de apendicitis practicadas durante un mes en hospitales generales del D.F., para lo cual se seleccionó una muestra de hospitales con probabilidad proporcional al número de camas por hospital y con reemplazo. El marco de muestreo estaba constituído por 1223 hospitales con un total de 63458 camas. Como el muestreo se realiza con reemplazo, algunos hospitales quedaron seleccionados más de una vez. A continuación se presentan los datos correspondientes por hospital: número de veces que quedó seleccionado, número de camas y número de operaciones.

Veces	Camas	Oper.	Veces	Camas	Oper.	Veces	Camas	Oper.
1	15	2	1	54	9	1	50	12
1	55	19	1	75	34	1	35	10
1	60	17	2	325	84	1	123	19
2	160	95	1	95	42	1	20	3
4	750	194	2	245	76	1	80	23
1	130	34	1	150	85	3	525	97
1	155	87	1	70	20	2	180	103
1	120	44	1	90	52	1	22	3
1	125	47	2	200	98	1	100	67
2	225	145	1	75	31	2	200	117

Cómputo:

Sección I

Como se trata de un problema de muestreo con probabilidad proporcional al tamaño, se elige [3] en I-1 (página 62) y aparece en pantalla lo siguiente:

```
1-MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO

2-MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON
  PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO Y EN LA
  SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO
!Numero que corresponde!
--> 1
```

```
*******************************************************************
* MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO *
*******************************************************************
1-Obtener un estimador
2-Determinar un tamano de muestra
3-Ambos
!Numero que corresponde!
--> 1

?Parametro a estimar?
1-Media
2-Total
!Numero que corresponde!
--> 2

?Total de unidades en la poblacion (N)?
--> 1223

?Tamano de muestra (n)?
--> 42

?Total de variable auxiliar? (X)
--> 63458

!Valores de interes (y) y Valores auxiliares (x)!
?Tiene los datos (y,x) en un archivo? (SI o NO)
--> SI

?Nombre de archivo = b:PPTEJE.DTA <--

?Quiere ver los datos? (SI o NO)
--> NO

?Quiere modificar algun dato? (SI o NO)
--> NO

?Quiere imprimir los datos? (SI o NO)
--> NO

?Quiere guardar los datos en un archivo? (SI o NO)
--> NO
```

Sección II

```
?Quiere imprimir los resultados?(SI o NO)
--> NO

** MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO **

** TOTAL **

* Estimador = 23652.2779

* Estimador de la varianza del estimador = 2724699.2053
```

```
Siendo estricto, se usa el percentil *Normal*
```

Nota: Esta decisión se toma con base en el tamaño de muestra que en este caso es 42.

```
    ?Esta de acuerdo? (SI o NO)
--> SI

    ?Con que confianza (%)?
--> 95

    * El intervalo de confianza al 95% para el total

    (    20416.9720 ,     26887.5838)  Z = 1.9600

    ?Quiere probar otro intervalo? (SI o NO)
--> NO

    *************Termino*************
    ?Quiere empezar de nuevo? (SI o NO)
--> NO

    ***************************
    *** Terminamos !GRACIAS! ***
    ***************************
```

4.6. EJEMPLO 6

Este ejemplo ilustra cómo estimar un total en un muestreo bietápico con selección en la primera etapa con probabilidad proporcional al tamaño con reemplazo y en la segunda etapa con muestreo aleatorio simple sin reemplazo.

Problema:

En una ciudad existen 150 supermercados distribuidos en 15 distritos. Los distritos no contiene necesariamente el mismo número de supermercados. Se desea estimar el total de latas de sopa que tienen en reserva los supermercados. La distribución de los supermercados se presenta a continuación:

Distrito	Número de supermercados
1	8
2	12
3	4
4	20
5	14
6	6
7	5
8	12
9	5
10	24
11	6
12	8
13	4
14	9
15	13

Se ha diseñado un muestreo en dos etapas utilizando como unidades primarias los distritos y como unidades secundarias los supermercados. Para la primera etapa se seleccionaron seis distritos con probabilidad proporcional al número de supermercados con reemplazo. En la segunda etapa se utilizó un muestreo aleatorio simple para seleccionar cuatro supermercados de cada distrito seleccionado, y en cada uno de ellos se determinó el número de latas en reseva.

En la selección de las unidades primarias los números aleatorios (entre 1 y 150) escogidos fueron 55, 148, 117, 70, 92 y 113. Entonces se seleccionaron los siguientes distritos:

distrito 5, 15, 12, 8, 10, 11.

Al realizar la segunda etapa de muestreo se obtuvieron los siguientes resultados:

	Números de latas en reserva			
distrito 5	2,	3,	2,	5
distrito15	2,	5,	5,	4
distrito12	1,	4,	8,	7
distrito 8	4,	8,	9,	7
distrito10	1,	1,	6,	4
distrito11	4,	3,	3,	6

(Trejo B. (1985) página 129)

Cómputo:

Sección I

Como se trata de un problema del muestreo bietápico con selección en la primera etapa con probabilidad proporcional al tamaño y en la segunda etapa con muestreo aleatorio simple sin reemplazo, se elige [3] en I-1 (página 62) y aparece en pantalla lo siguiente:

```
    1-MUESTREO CON PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO

    2-MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON
      PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO Y EN LA
      SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO
    !Numero que corresponde!
--> 2
```

```
    *****************************************************************
    *** MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON  ***
    *** PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO Y EN LA ***
    *** SUGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO ***
    *****************************************************************
    ?Parametro a estimar?
    1-Media
    2-Total
    !Numero que corresponde!
--> 2
```

```
    ?Total de unidades primarias (N)?
--> 15
```

```
    ?Total de unidades primarias en la muestra (n)?
--> 6
```

```
    ?Total de unidades secundarias en la poblacion (M)?
--> 150
```

```
    ?Total de unidades secundarias en cada unidad primaria en la
    muestra (Mi)?
    1-esima unidad primaria = 14
    2-esima unidad primaria = 13
    3-esima unidad primaria = 8    <--
    4-esima unidad primaria = 12
    5-esima unidad primaria = 24
    6-esima unidad primaria = 6
```

```
    ?Quiere ver los datos? (SI o NO)
--> NO
```

```
    ?Quiere modificar algun dato? (SI o NO)
--> NO
```

```
?Total de unidades secundarias muestreadas en cada unidad primaria
(mi)?
1-esima unidad primaria = 4
2-esima unidad primaria = 4
3-esima unidad primaria = 4   <--
4-esima unidad primaria = 4
5-esima unidad primaria = 4
6-esima unidad primaria = 4
```

```
Total de unidades secundarias en la muestra =  24
```

```
    ?Quiere ver los datos? (SI o NO)
--> NO
```

```
    ?Quiere modificar algun dato? (SI o NO)
--> NO
```

```
    !Valores de interes (y)!
    ?Tiene los datos (y) en un archivo? (SI o NO)
--> NO
```

```
!Introduzca los datos (y) como se pide!
** 1-esima unidad primaria **
y[ 1, 1] = 2
y[ 1, 2] = 3    <--
y[ 1, 3] = 2
y[ 1, 4] = 5
** 2-esima unidad primaria **
y[ 2, 1] = 2
y[ 2, 2] = 5    <--
y[ 2, 3] = 5
y[ 2, 4] = 4
** 3-esima unidad primaria **
y[ 3, 1] = 1
y[ 3, 2] = 4    <--
y[ 3, 3] = 8
y[ 3, 4] = 7
** 4-esima unidad primaria **
y[ 4, 1] = 4
y[ 4, 2] = 8    <--
y[ 4, 3] = 9
y[ 4, 4] = 7
** 5-esima unidad primaria **
y[ 5, 1] = 1
y[ 5, 2] = 1    <--
y[ 5, 3] = 6
y[ 5, 4] = 4
** 6-esima unidad primaria **
y[ 6, 1] = 4
y[ 6, 2] = 3    <--
y[ 6, 3] = 3
y[ 6, 4] = 6
```

Nota: Debe observarse que el orden en el que se introducen los datos de las unidades primarias (el número total de unidades secundarias, el tamaño de muestra y la información muestral) debe ser siempre mismo.

```
    ?Quiere ver los datos? (SI o NO)
--> NO

    ?Quiere modificar algun dato? (SI o NO)
--> NO

    ?Quiere imprimir los datos? (SI o NO)
--> NO

    ?Quiere guardar los datos en un archivo? (SI o NO)
--> SI

    ?Nombre de archivo = b:bietapic.dta <--
```

Nota: Al dar el nombre del archivo los datos quedan grabados en el disco de la UNIDAD-B.

Sección II

*** RESULTADOS PARCIALES DE CADA UNIDAD PRIMARIA ***

i *	1*	2*	3*	4*	5*
M *	14*	13*	8*	12*	24*
m *	4*	4*	4*	4*	4*
My *	3.00*	4.00*	5.00*	7.00*	3.00*
Ty *	42.00*	52.00*	40.00*	84.00*	72.00*

|Introduzca 'RETURN' _

i *	6*
M *	6*
m *	4*
My *	4.00*
Ty *	24.00*

|Introduzca 'RETURN' _

```
    ?Quiere imprimir los resultados parciales? (SI o NO)
--> NO
```

```
?Quiere imprimir los resultados? (SI o NO)
--> NO
```

```
*** MUESTREO BIETAPICO CON SELECCION EN LA PRIMERA ETAPA CON  ***
*** PROBABILIDAD PROPORCIONAL AL TAMANO CON REEMPLAZO Y EN LA ***
*** SEGUNDA ETAPA CON MUESTREO ALEATORIO SIMPLE SIN REEMPLAZO ***

** TOTAL **

* Estimador = 650.0000

* Estimador de la varianza del estimador = 8500.0000
```

```
?La caracteristica tiene distribucion normal? (SI o NO)
--> NO
```

```
Siendo estricto, se usa el percentil *Tchebycheff*
```

Nota: Aparece este mensaje en pantalla tomando en cuenta las condiciones para decidir qué percentil se debe utilizar (página 59).

```
?Esta de acuerdo? (SI o NO)
--> SI
```

```
?Con que confianza? (%)
--> 95
```

```
* El intervalo de confianza al 95% para el total

 (       237.6894 ,       1062.3106 )    K = 4.4721
```

```
?Quiere probar otro intervalo? (SI o NO)
--> NO
```

```
*************Termino*************
?Quiere empezar de nuevo? (SI o NO)
--> NO
```

```
****************************
*** Terminamos !GRACIAS! ***
****************************
```

BIBLIOGRAFIA

- Cochran G. William, (1974), "Técnicas de Muestreo"
  John Wiley & Sons, Inc. N.Y
- Des RaJ, (1968), "Sampling Theory"
  McGRAW HILL N.Y
- Keller M. Arthur, (1982), "Programación en PASCAL"
  McGROW HILL N.Y
- Mendenhall William, Ott Lynan, Sheaffer L Richard, (1986),
  "Elementos de Muestreo" W.P.S Inc.
- Mendez Ignacio, (1976), "Concepto muy elemental del muestreo
  con énfasis en la determinación práctica del tamaño de muestra"
  Comunicaciones Técnicas, Serie azul:Monografía No.25
  Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
  (I.I.M.A.S) U.N.A.M
- P.V. Sukhatame & B.V. Sukhatame
  "Sampling Theory of Survey with Applications" 1970
  Iowa State University Press
- TreJo V. Belem, (1985), "Introducción a técnicas de muestreo:
  Ejercicios de aplicación de temas de un curso básico"
  Comunicacion Técnicas Serie verde: Notas No.29
  I.I.M.A.S U.N.A.M
- Manual Microsoft MS-DOS Version 2.0
  Operating system User's Guide 1983
- Manual Turbo Pascal Version 3.0
  Reference Manual 1985
  Borland International Inc.