115

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

LA EFICIENCIA DE ALGUNOS ESQUEMAS MUESTRALES APLICADOS A DATOS SIMULADOS CON DIFERENTES GRADOS DE ASOCIACION

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A

P R E S E N T A :

NANCY BEATRIZ ZUÑIGA HIDALGO

DIRECTOR DE TESIS:

M. EN C. JOSE ANTONIO FLORES DIAZ

1998

VNIVERSIDAD NACIONAL
AVFNºMA DE
MEXICO

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

Comunicamos a usted que hemos revisado el trabajo de Tesis:
La eficiencia de algunos esquemas muestrales aplicados a datos simulados con diferentes grados de asociación.

realizado por Nancy Beatriz Zúñiga Hidalgo

con número de cuenta 8723811-6 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario M. en C. José Antonio Flores Díaz

Propietario M. en A. P. María del Pilar Alonso Reyes

Propietario Act. María Guadalupe Tzintzún Cervantes

Suplente Act. Yazmín Iliana Bárcenas Orozco

Suplente Act. Tania Chávez Razo

Consejo Departamental de Matemáticas
M. en A. P. María del Pilar Alonso Reyes

A mis padres. Gracias por brindarme su apoyo, confianza y amor.

A mi hermano y hermanas.

Quiero expresar un sincero agradecimiento a mis sinodales, quienes participaron en la revisión de este trabajo:

M. en A. P. María del Pilar Alonso Reyes
Act. María Guadalupe Tzintzún Cervantes
Act. Yazmín Iliana Bárcenas Orozco
Act. Tania Chávez Razo

Un agradecimiento especial para mi director de tesis M. en C. José Antonio Flores Díaz, por su apoyo, paciencia, amistad, etc.

Gracias. A todos mis amigos por formar parte de mi vida.

Claudia P., Claudia S., Mary, Josefina, Karla, Pilar,

Rita, Yadira, Ricardo, Osvaldo, Luis Felipe, etc.

Gracias. A la UNAM por brindarme la oportunidad de estudiar, y permitirme el uso de su equipo (CRAY) para la elaboración del presente trabajo.

Indíce

INTRODUCCIÓN

Para obtener información sobre fenómenos físicos, biológicos, sociales, artes, ciencias, etc., el investigador recopila información a través de la experimentación o almacena datos que representan el comportamiento del fenómeno a estudiar en un período determinado de tiempo, a estos datos se les aplican técnicas estadísticas para obtener información (media, medidas de dispersión, etc.) que facilitan el análisis e interpretación de dicho fenómeno.

Pero en ocasiones sólo se tiene parte de los datos del fenómeno o simplemente son demasiados para trabajar con ellos, ya que esto acarrearía problemas de tiempo, costo, equipo, etc., entonces el investigador recurre a hacer estimaciones que le facilitan el estudio de dicho fenómeno.

Para realizar estas estimaciones se han desarrollado varias teorías y técnicas (muestreo aleatorio simple, muestreo aleatorio con probabilidad proporcional al tamaño de las unidades, etc.), ahora el problema es el saber cual es el mejor estimador.

Una forma para determinar cual es el mejor estimador con respecto a otro es a través de su variabilidad, esto es, teóricamente, un cierto estimador puede resultar tener una menor variabilidad que algún otro, pero esta variabilidad numéricamente puede ser insignificante por esta razón, Flores (1992) desarrolló un análisis teórico y numérico de la variabilidad de la media poblacional para el muestreo aleatorio simple, en el cual se revisaron los estimadores directo, razón y regresión, para el muestreo aleatorio con probabilidad proporcional al tamaño de las unidades se consideraron los estimadores con y sin (modelo de Des Raj o secuencial) reemplazo, en donde se estudiaron 17 modelos o tipos de relación entre la variable independiente y dependiente, las expresiones en cuestión, contempladas abarcan tanto funciones lineales (6 funciones), cuadráticas (3), exponenciales (4) y logarítmicas (4), en el que concluyó que el mejor estimador fue el de regresión, pero las funciones estudiadas alcanzaron niveles de "perfecta correlación", esto es 99.997 ó 100 ó (-100) por ciento.

Por consiguiente el presente trabajo hace un análisis teórico y numérico de la variabilidad de la media poblacional para el muestreo aleatorio simple para los estimadores directo, razón y regresión, para el muestreo aleatorio con probabilidad proporcional al tamaño de las unidades para los estimadores de con y sin (Des Raj) reemplazo, en el cual se estudian valores con la variable auxiliar X y valores de la característica en estudio Y relacionados con distintos grados de correlación (cov = 0.3, 0.5 y 0.8).

Para realizar dicho estudio se realiza una simulación en la computadora CRAY de la Dirección General de Servicios de Computo Académico de la UNAM.

En el capítulo uno se presentan los resultados teóricos del muestreo aleatorio simple sin reemplazo para los estimadores directo, razón y regresión de la media.

En el capítulo dos se presentan los resultados teóricos del muestreo aleatorio con probabilidad proporcional al tamaño de las unidades para los estimadores de con y sin reemplazo de la media.

El capítulo tres presenta comentarios relacionados con los elementos o condiciones con base a los cuales se desarrolló la simulación, también se refieren los archivos de información generados por los programas.

En el capítulo cuatro se realizan comentarios correspondientes a las corridas de los programas de la simulación, así como también a las de los programas desarrollados en SPSSX, para obtener resultados con relación a los datos producidos por los programas anteriores.

En el capítulo cinco se realiza una síntesis de los resultados obtenidos de la simulación.

En el apéndice A se presentan algunas demostraciones de resultados utilizados en el capítulo uno.

En el apéndice B se presentan los resultados obtenidos subdivididos de la siguiente manera:

B.1. Cuadros de resultados concentrados para medias y varianzas muestrales.

B.2. Cuadros de estadísticas calculadas para la prueba de Kolmogorov-Smirnov por modelo.

En el apéndice C se exhiben las subrutinas que sufrieron modificaciones en el programa que simuló el muestreo aleatorio simple sin reemplazo, desarrollando los estimadores directo, de razón y de regresión, en este caso, además las rutinas con modificaciones del programa que simuló el muestreo aleatorio con probabilidad proporcional al tamaño de las unidades, contemplando los estimadores de Des Raj o secuencial y uno para la media, en los casos de sin y con reemplazo respectivamente; ambos programas se desarrollaron en FORTRAN 77. Igualmente se muestran los cuatro programas de SPSSX que se emplearon para analizar la información producto de las simulaciones realizadas.

Capítulo I

RESULTADOS DEL MUESTREO ALEATORIO SIMPLE

I.1.- Introducción

El análisis estadístico inferencial se refiere al conjunto de teorías y técnicas que conducen a explicar el comportamiento de fenómenos de naturaleza aleatoria, con base a la información disponible acerca de ellos. Los procedimientos en cuestión trabajan sobre un subconjunto de elementos de la población, comúnmente llamada muestra, que se considera representativo de ella y cuyos elementos son seleccionados aleatoriamente; obteniendo, de ellos, resultados que se generalizan al resto de la población.

De lo anterior, se desprenden dos conceptos fundamentales a saber:

Población, que es la colección de elementos, ya sea que éstos sean objetos, seres vivos, hechos concretos, etc., de los cuales interesa toda información que los caracteriza como fenómeno, es decir, el total de "cosas" que se han escogido como objeto de estudio.

Muestra, que es el subconjunto de elementos de la población, elegidos por un algoritmo o procedimiento previamente planeado, que tiene como propósito caracterizar a los componentes de él con las cualidades de **aleatoriedad** y **representatividad** y que sólo dicho algoritmo es el que permite cualificar de esta manera a la muestra.

En el desarrollo y concretización del análisis estadístico inferencial son muchos los aspectos que se han considerado, en particular uno más que resulta importante delimitar en la medida de lo posible, es el que corresponde a la eficiencia, en términos del error cuadrático medio e insesgamiento, que un estimador puede ofrecer, dependiendo, si es el caso, del esquema o diseño de muestreo.

En las siguientes secciones se revisará los resultados básicos que tiene que ver con el muestreo aleatorio simple.

I.2 .- Muestreo Aleatorio Simple

El muestreo aleatorio simple es un método de selección de n unidades en un conjunto de N de tal modo que cada una de la $\frac{N!}{n!(N-n)!}$ muestras distintas tengan la misma oportunidad de ser elegidas.

La teoría de muestreo es útil para calcular ciertas características de la muestra aleatoria y con base en ellas la estimación de la media poblacional; este proceso se puede desarrollar a través de tres técnicas, esquemas o enfoques de calculo:

i) Directo o simple

ii) De razón

iii) De regresión

Para efectos de desarrollo de las técnicas antes mencionadas se utilizará la siguiente notación:

Las letras mayúsculas señalan a las características de la población y las minúsculas a las de la muestra. El símbolo ^ identifica una estimación muestral de una característica de la población.

Y_i Valor de la característica en estudio para el individuo *i* de la población.

N Número total de elementos o individuos de la población.

y_i Valor de la característica en estudio para el individuo *i* de la muestra

n Número total de elementos o individuos seleccionados para formar la muestra.

$\overline{Y}$ Valor promedio de la característica en estudio en la población (media poblacional).

$\bar{y}$ Valor promedio de la característica en estudio en la muestra (media muestral).

$\hat{\overline{Y}}$ Estimador de la media poblacional.

Y Valor total de las características de estudio de la población (total poblacional).

y Valor del total de la característica de estudio de la muestra.

$\hat{Y}$ Estimador del total de las características de estudio de la población (estimador del total poblacional).

R Razón poblacional con relación a dos características en estudio.

$\hat{R}$ Estimador de la razón poblacional.

C_n^N Total de combinaciones tomadas de n en n de N.

i) El estimador directo que determina la media poblacional está dado como:

$$\hat{\bar{Y}} = \bar{y} = \sum_{i=1}^{n} \frac{y_i}{n} \tag{I.1}$$

que es un estimador insesgado[1]

La varianza de la media poblacional correspondiente está determinada por:

$$Var(\hat{\bar{Y}}) = (1 - \frac{n}{N}) \frac{S_y^2}{n} \tag{I.2}$$

con

$$S_y^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N - 1} \tag{I.3}$$

ii) Con respecto al estimador de razón, la media poblacional se plantea como:

$$\hat{\bar{Y}}_{Ra} = \frac{\bar{y}}{\bar{x}} \bar{X} = \frac{y}{x} \bar{X} \tag{I.4}$$

[1] Mendenhall, *Estadística matemática con aplicaciones*, ver capítulo 8, definición 8.2, pág. 300.

es un estimador sesgado, por lo tanto se considera al error cuadrático medio[2] (ECM) como medida de variabilidad en lugar de la varianza.

$$ECM(\hat{\bar{Y}}_{Ra}) \cong \frac{(1-\frac{n}{N})}{n}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right] \tag{I.5}$$

con S_y^2 definida por la expresión (I.3) y S_x^2 es idéntica, sólo que se sustituye las y por las x y

$$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y}) \tag{I.6}$$

con

$$R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}} \tag{I.7}$$

iii) En cuanto al estimador de regresión, su media poblacional es estimada por

$$\hat{\bar{Y}}_{Re} = \bar{y} + \hat{B}(\bar{X} - \bar{x}) = \bar{y} + b(\bar{X} - \bar{x}) \tag{I.8}$$

donde $\hat{B}$ o b está definida por:

$$\hat{B} = b = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{I.9}$$

[2] Mood, *Introduction the Thery of Statistics*, 1974 ver capítulo VII, definición 6, pág. 291.

En este caso, también se considerara al error cuadrático medio como medida de variabilidad dado que el valor esperado del estimador es sesgado, la fórmula correspondiente es:

$$ECM(\hat{\bar{Y}}_{Rc}) \cong \frac{(1-\frac{n}{N})}{n} S_y^2(1-\rho^2) \tag{I.10}$$

con

$$\rho = \frac{\dfrac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{N-1}}{\sqrt{\dfrac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}}\sqrt{\dfrac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}}}$$

Las expresiones (I.1), (I.4) y (I.8) dan una aproximación al valor de la media poblacional de la característica en estudio, aunque estos resultados se pueden aplicar para obtener el total; sin embargo en el presente trabajo, sólo se hablará de la media poblacional.

Las expresiones (I.2), (I.5) y (I.10) proporcionan la variabilidad o el error cuadrático medio correspondiente a cada uno de los estimadores en estudio. No obstante dichos estimadores nunca son conocidos ya que dependen de elementos poblacionales desconocidos, tales como S_y^2, R y ρ^2, respectivamente, por esta razón, es necesario hacer aproximaciones, dando como resultado las siguientes expresiones:

Para la expresión (I.2), el estimador correspondiente sería:

$$\hat{Var}(\hat{\bar{Y}}) = \text{var}(\hat{\bar{Y}}) = (1-\frac{n}{N})\frac{s_y^2}{n} \tag{I.11}$$

con

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} \tag{I.12}$$

En el caso de la expresión (I.5) la estimación se da como

$$\hat{E}CM\,(\hat{\bar{Y}}_{Ra}) = ecm\,(\hat{\bar{Y}}_{Ra})$$

$$\cong \frac{(1-\frac{n}{N})}{n}(s_y^2 + r^2 s_x^2 - 2rs_{xy}) \tag{I.13}$$

con s_y^2 igual a la expresión (I.12), s_x^2 idéntica a s_y^2 , sólo se sustituyen las y por las x y s_{xy} dada por

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{I.14}$$

y

$$\hat{R} = r = \frac{y}{x} = \frac{\bar{y}}{\bar{x}} \tag{I.15}$$

Por último, con respecto a la expresión (I.10), la estimación en cuestión se efectúa de la siguiente manera

$$ecm(\hat{\bar{Y}}_{Re}) \cong \frac{(1-\frac{n}{N})}{n} s_y^2(1-\hat{\rho}^2) \tag{I.16}$$

con s_y^2 igual a la expresión (I.12) y

$$\hat{\rho} = \frac{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}} \qquad \text{(I.17)}$$

I.3 .- Los Valores Esperados de los Estimadores Considerados

Los valores esperados de los estimadores ya enunciados son:

En el caso del estimador directo o simple, se tiene que los valores esperados de las ecuaciones (I.1) y (I.11) son respectivamente:

$$E\left[\hat{\bar{Y}}\right] = E[\bar{y}] = E\left[\sum_{i=1}^{n}\frac{y_i}{n}\right]$$

$$= \frac{1}{n}\frac{C_{n-1}^{N-1}}{C_n^N}\sum_{i=1}^{N} Y_i$$

$$= \frac{(N-1)!\,n!\,(N-n)!}{nN!\,(n-1)!\,(N-n)!}\sum_{i=1}^{N} Y_i$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i = \bar{Y} \qquad \text{(I.18)}$$

y

$$E\left[\mathrm{var}(\hat{\bar{Y}})\right] = E\left[(1-\frac{n}{N})\frac{s_y^2}{n}\right]$$

$$= (1-\frac{n}{N})\frac{E\left[s_y^2\right]}{n}$$, utilizando el resultado (A.2) que se

demuestra en el apéndice A se obtiene:

$$E\left[\text{var}(\hat{\bar{Y}})\right] = (1-\frac{n}{N})\frac{S_y^2}{n} \qquad (I.19)$$

siendo estimadores insesgados.

Con relación al estimador de razón, la expresión (I.4) tiene como valor esperado

$$E\left[\hat{\bar{Y}}_{Ra}\right] = E\left[\hat{R}\bar{X}\right]$$

aplicando el teorema[3] 5.5 y utilizando los resultados (A.5) y (A.9) del apéndice A, se obtiene:

$$E\left[\hat{\bar{Y}}_{Ra}\right] = \bar{X}E\left[\hat{R}\right] = \bar{X}E\left[\frac{\bar{y}}{\bar{x}}\right]$$

$$\cong \bar{X}R = \bar{Y}$$

Esto justifica el por que el error cuadrático medio se utiliza como medida de variabilidad en lugar de la varianza.

El error cuadrático medio está dado por:

$$E\left[(\hat{R}\bar{X} - \bar{Y})^2\right]$$

[3] Mendenhall et al, 1991 ver capítulo 5, teorema 5.5, pág. 203

aplicando el teorema[4] de Taylor y utilizando el resultados (A.10) del apéndice A se obtiene:

$$ECM(\hat{\bar{Y}}_{Ra}) = E\left[(\hat{R}\bar{X} - \bar{Y})^2\right]$$

$$= E\left[\frac{\bar{X}^2(\bar{y} - R\bar{x})^2}{\bar{x}^2}\right]$$

$$\cong \bar{X}^2\left[E\left[\frac{(\bar{y} - R\bar{x})^2}{\bar{X}^2}\right] - \sum_{i=1}^{\infty} E\left[\frac{(-1)^i(\bar{y} - R\bar{x})^2(\bar{x}^2 - \bar{X}^2)^i}{(\bar{X}^2)^{i+1}}\right]\right] \qquad (I.20)$$

Se consideran como insignificantes las expresiones del segundo término en adelante, del lado derecho de la igualdad (I.20), ya que el denominador cada vez es más grande que el numerador por su potencia, en consecuencia el error cuadrático medio de $\hat{\bar{Y}}_{Ra}$, estaría tan sólo determinado por $E\left[(\bar{y} - R\bar{x})^2\right]$, por ello, se dice que

$$ECM(\hat{\bar{Y}}_{Ra}) = E\left[(\hat{\bar{Y}}_{Ra} - \bar{Y})^2\right] = E\left[(\frac{\bar{y}}{\bar{x}}\bar{X} - \bar{Y})^2\right]$$

$$= E\left[(\frac{\bar{y}\bar{X} - \bar{Y}\bar{x}}{\bar{x}})^2\right] \cong E\left[(\frac{\bar{y}\bar{X} - \bar{Y}\bar{x}}{\bar{X}})^2\right]$$

$$= E\left[(\bar{y} - \frac{\bar{Y}}{\bar{X}}\bar{x})^2\right] = E\left[(\bar{y} - R\bar{x})^2\right]$$

aplicando el resultado (A.15) del apéndice A se obtiene:

[4] Mood et al, 1974 ver capítulo V, teorema 4, pág. 181

$$ECM(\hat{\bar{Y}}_{Ra}) = \frac{(1-\frac{n}{N})}{n}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right] \tag{I.21}$$

Por consiguiente la expresión (I.13) sólo podría ser un estimador insesgado de la última parte de la expresión (I.21), sin embargo resulta que $s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}$ es equivalente a

$$\sum_{i=1}^{n} \frac{(y_i - \hat{R}x_i)^2}{n-1} \tag{I.22}$$

el cual es un estimador muestral que representa un sesgo de orden 1/n [5], cuyo principal término en el sesgo esta dado por la expresión (I.13) que a su vez es una estimación sesgada de (I.5).

Por otra parte, en el caso del estimador de regresión se observa que (I.8) es una estimación con un sesgo de 1/n [6], cuyo principal término en el sesgo está dado por la expresión:

$$-\frac{(1-\frac{n}{N})}{n}\frac{E\left[e_i(x_i - \bar{X})^2\right]}{S_x^2} \tag{I.23}$$

con

$$e_i = y_i - \bar{Y} - B(x_i - \bar{X}) \tag{I.24}$$

que tiene las propiedades:

$$\sum_{i=1}^{N} e_i = 0$$

[5] Cochran, 1992, ver sección 2.11 pág. 56-60
[6] Ibid sección 7.7. págs. 249-250

$$\sum_{i=1}^{N} e_i(x_i - \overline{X}) = \sum_{i=1}^{N} (y_i - \overline{Y})(x_i - \overline{X}) - B\sum_{i=1}^{N} (x_i - \overline{X})^2 = 0$$

por lo tanto, la esperanza correspondiente utilizando el resultado (A.17) del apéndice A es:

$$E\left[\hat{\overline{Y}}_{Re}\right] = \overline{Y} - E\left[b(\bar{x} - \overline{X})\right] = \overline{Y} - Cov(b, \bar{x})$$

aplicando el resultado (A.23) del apéndice A

$$E\left[\hat{\overline{Y}}_{Re}\right] \cong \overline{Y} \qquad (I.25)$$

en consecuencia, al igual que en el caso del estimador de razón, se nota que únicamente es posible referir el concepto de error cuadrático medio, cuya estimación (I.16) resulta ser igualmente sesgada[7], por un término cuyo orden es $1/\sqrt{n}$, dando por consiguiente un estimador sesgado del lado derecho de (I.10).

I.4 .- Análisis Comparativo de los Errores Cuadráticos Medios

El análisis comparativo de los valores esperados respecto a los estimadores de la media poblacional, hasta ahora efectuado, no permite decir cual es el mejor estimador en cuestión y cual es el menos recomendable; por ello, es necesario hacer un análisis profundo, desde el punto de vista estadístico, que proporcione elementos de decisión más convincentes.

Por esta razón, conjuntamos tanto el análisis de los valores esperados de los estimadores, como el de los errores cuadráticos medios, los cuales se mencionan a continuación.

[7] Cochran, Sección 7.4 pág. 245

En muestras grandes y considerando un muestreo aleatorio simple, el estimador de razón para la media poblacional, tiene un error cuadrático medio menor, que el correspondiente al estimador directo o simple si

$$\rho > \frac{1}{2}\frac{\frac{S_x}{\bar{\bar{X}}}}{\frac{S_y}{\bar{\bar{Y}}}}$$

$$ECM(\bar{Y}_S) = Var(\bar{Y}_S) = \frac{(1-\frac{n}{N})}{n}S_y^2$$

$$ECM(\hat{\bar{\bar{Y}}}_{Ra}) \cong \frac{(1-\frac{n}{N})}{n}\cdot(S_y^2 + R^2S_x^2 - 2RS_{xy})$$

$$\Rightarrow \frac{(1-\frac{n}{N})}{n}S_y^2 > \frac{(1-\frac{n}{N})}{n}\cdot(S_y^2 + R^2S_x^2 - 2RS_{xy})$$

$$S_y^2 > S_y^2 + R^2S_x^2 - 2RS_{xy}$$

$$0 > R^2S_x^2 - 2RS_{xy}$$

$$2RS_{xy} > R^2S_x^2$$

$$S_{xy} > \frac{RS_x^2}{2}$$

$$\frac{S_{xy}}{S_xS_y} > \frac{RS_x^2}{2S_xS_y}$$

$$\rho > \frac{RS_x}{2S_y}$$

$$\rho > \frac{\frac{\overline{Y}}{\overline{X}}S_x}{2S_y}$$

$$\rho > \frac{S_x\overline{Y}}{2S_y\overline{X}} \qquad \text{(I.26)}$$

En el caso del estimador directo o simple de la media poblacional, su error cuadrático medio es igual a la varianza de dicho estimador, por ser este un estimador insesgado.

Con relación al estimador de regresión, se observa de (I.2) y (I.10), que sólo cuando la correlación entre las variables X y Y es nula, los errores cuadráticos medios correspondientes son iguales, es decir

$$Var(\hat{\overline{Y}}_S) = \frac{S_y^2}{n}(1-\frac{n}{N})$$

$$ecm(\hat{\overline{Y}}_{Re}) \cong \frac{(1-\frac{n}{N})}{n}S_y^2(1-\rho^2)$$

$$\frac{S_y^2}{n}(1-\frac{n}{N}) = (1-\frac{n}{N})\frac{S_y^2}{n}(1-\rho^2)$$

si y solo si $\rho = 0$, entonces, a medida que la correlación aumenta, en valor absoluto, el error cuadrático medio del estimador de regresión es mejor que el estimador simple.

Considerando los errores cuadráticos medios de los estimadores de razón y de regresión de las expresiones (I.5) y (I.8), se observa que:

$$\frac{(1-\frac{n}{N})}{n}(S_y^2 + R^2S_x^2 - 2RS_{xy}) > \frac{(1-\frac{n}{N})}{n}S_y^2(1-\rho^2)$$

$$S_y^2 + R^2S_x^2 - 2RS_{xy} > S_y^2(1-\rho^2)$$

$$R^2S_x^2 - 2RS_{xy} > -S_y^2\rho^2 \qquad (I.27)$$

lo cual equivale a

$$-\rho^2S_y^2 < R^2S_x^2 - \frac{2S_{xy}}{S_xS_y}RS_xS_y$$

$$-\rho^2S_y^2 < R^2S_x^2 - 2\rho RS_xS_y \qquad (I.28)$$

que a su vez puede considerarse como

$$-\rho^2S_y^2 + 2\rho RS_xS_y - R^2S_x^2 = -(\rho S_y - RS_x)^2 < 0 \qquad (I.29)$$

o bien $$(\rho S_y - RS_x)^2 > 0 \qquad (I.30)$$

finalmente, factorizando S_x^2, se obtiene:

$$(\frac{\rho S_y}{S_x} - R)^2 = (B-R)^2 > 0 \qquad (I.31)$$

con base a lo anterior, se concluye que el estimador de regresión es mejor, en el sentido de un error cuadrático medio mas pequeño, que el estimador de razón; así mismo sus respectivos errores cuadráticos medios serán iguales si y solo si $B = R$. Lo cual equivale a especificar que la relación que existe entre X y Y es una recta que pasa por el origen.

RESULTADOS DEL MUESTREO CON PROBABILIDAD PROPORCIONAL AL "TAMAÑO".

II.1.- Introducción

En el presente capítulo se aborda el tema del muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades, considerando las opciones de con y sin reemplazo. Esta técnica generalmente es usada para la selección de unidades grandes así como en el caso de ciudades, villas o manzanas etc; Cabe señalar que el término "tamaño" se indica entre comillas, ya que, puede, indistintamente, referirse a diversas características, dependiendo de la población y situación particular de aplicación o desarrollo del muestreo en cuestión.

En el muestreo aleatorio con probabilidades desiguales y sin reemplazo, las expresiones de la variabilidad del estimador, para la media o el total poblacional, involucran el cálculo de la probabilidad de que dos unidades cualesquiera de la población sean seleccionadas conjuntamente, situación que resulta casi siempre compleja.

Así mismo, este tipo de muestreo presenta, teóricamente, la posibilidad, bajo ciertas condiciones[1], de tener una varianza más pequeña, que la correspondiente a un muestreo aleatorio simple con reemplazo.

Con base, en las consideraciones anteriores, se determinó estudiar el muestreo con probabilidad proporcional al "tamaño" para dos opciones o procesos de estimación menos difíciles de implantar o desarrollar, como son: el estimador secuencial o de Des Raj en el caso de sin reemplazo, y un estimador de la media para el de con reemplazo; los cuales se desarrollan en este capítulo.

[1] Raj, 1968, corolario pág. 49 y sección 3.15 pág. 50.

II.2 .-Muestreo Aleatorio con probabilidad Proporcional al "Tamaño" con Reemplazo.

II.2.1 Especificación del Proceso de Estimación.

La teoria del muestreo propone como un estimador del total poblacional a

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{P_i} \tag{II.1}$$

donde

$$P_i = \frac{X_i}{X} \quad i = 1, 2, 3, \ldots .N$$

P_i representa la probabilidad de que el elemento i de la población sea seleccionado y X_i es el valor de la característica considerada como el "tamaño" de dicho elemento, mientras que el total para la población está dado por

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

la variabilidad muestral de este estimador es

$$Var_1(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{N} P_i\left(\frac{Y_i}{P_i} - Y\right)^2 \tag{II.2}$$

que alternativamente puede expresarse como

$$Var_2(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\left[P_i P_j\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2\right] \tag{II.3}$$

De igual forma que en el capítulo anterior, las expresiones (II.2) y (II.3) tienen que estimarse, debido a que su determinación implica a todos los elementos de la población, lo cual no es posible ni lógico; en consecuencia, los estimadores de dichas expresiones son:

$$\operatorname{var}_1(\hat{Y}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{P_i} - \hat{Y}\right)^2 \qquad \text{(II.4)}$$

$$\operatorname{var}_2(\hat{Y}) = \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left[\left(\frac{y_i}{P_i} - \frac{y_j}{P_j}\right)^2\right] \qquad \text{(II.5)}$$

II.2.2 Los Valores Esperados de los Estimadores

No es difícil verificar que

$$E\left[\hat{Y}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{P_i}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\frac{y_i}{P_i}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}E\left[\frac{y_i}{P_i}\right] = \frac{n}{n}E\left[\frac{y_i}{P_i}\right]$$

$$= \sum_{i=1}^{N}\frac{Y_i P_i}{P_i} = \sum_{i=1}^{N}Y_i = Y \qquad \text{(II.6)}$$

es decir, que (II.1) es un estimador insesgado del total poblacional.

Así mismo se puede demostrar que (II.4) y/o (II.5) son estimadores insesgados de (II.2) y/o (II.3), en virtud de que estas últimas, son expresiones iguales, por lo tanto se tiene que:

$$E\left[\operatorname{var}_1(\hat{Y})\right] = E\left[\operatorname{var}_2(\hat{Y})\right] = Var_1(\hat{Y}) = Var_2(\hat{Y}) \qquad \text{(II.7)}$$

En el caso de la media poblacional, estos resultados se pueden estimar de la siguiente manera:

$$\hat{\bar{Y}} = \frac{1}{N}\hat{Y} = \frac{1}{nN}\sum_{i=1}^{n}\frac{y_i}{P_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{NP_i} \tag{II.8}$$

$$Var_1(\hat{\bar{Y}}) = \frac{1}{N^2}Var_1(\hat{Y}) = \frac{1}{nN^2}\sum_{i=1}^{N}P_i\left(\frac{Y_i}{P_i} - Y\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{N}P_i\left(\frac{Y_i}{NP_i} - \bar{Y}\right)^2 \tag{II.9}$$

en cuanto a la expresión (II.3)

$$Var_2(\hat{\bar{Y}}) = Var_2\left(\frac{1}{N}\hat{Y}\right) = \frac{1}{N^2}Var_2(\hat{Y})$$

$$\frac{Var_2(\hat{Y})}{N^2} = \frac{1}{N^2 n}\sum_{i=1}^{N}\sum_{j>i}^{N}\left[P_iP_j\left(\frac{Y_i}{P_i} - \frac{Y_j}{P_j}\right)^2\right] = \frac{1}{n}\sum_{i=0}^{N}\sum_{j>i}^{N}\left[P_iP_j\left(\frac{Y_i}{NP_i} - \frac{Y_j}{NP_j}\right)^2\right] \tag{II.10}$$

por lo que corresponde a los estimadores de (II.9) y (II.10), se puede mostrar que son

$$\mathrm{var}_1(\hat{\bar{Y}}) = \mathrm{var}_1\left(\frac{\hat{Y}}{N}\right)$$

$$= \frac{\mathrm{var}_1(\hat{Y})}{N^2}$$

$$= \frac{1}{N^2 n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{P_i} - \hat{Y}\right)^2$$

/

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \frac{1}{N^2} \left(\frac{y_i}{P_i} - \hat{Y}\right)^2$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{NP_i} - \frac{\hat{Y}}{N}\right)^2$$

$$\mathrm{var}_1(\hat{\bar{Y}}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{NP_i} - \hat{\bar{Y}}\right)^2 \tag{II.11}$$

y

$$\mathrm{var}_2(\hat{\bar{Y}}) = \mathrm{var}_2\left(\frac{\hat{Y}}{N}\right)$$

$$= \frac{\mathrm{var}_2(\hat{Y})}{N^2}$$

$$= \frac{1}{N^2 n^2 (n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \left(\frac{y_i}{P_i} - \frac{y_j}{P_j}\right)^2$$

$$= \frac{1}{n^2 (n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{1}{N^2} \left(\frac{y_i}{P_i} - \frac{y_j}{P_j}\right)^2$$

$$= \frac{1}{n^2 (n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \left(\frac{y_i}{NP_i} - \frac{y_j}{NP_j}\right)^2$$

$$\mathrm{var}_2(\hat{\bar{Y}}) = \frac{1}{n^2 (n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \left(\frac{y_i}{NP_i} - \frac{y_j}{NP_j}\right)^2 \tag{II.12}$$

II.3.- Muestreo Aleatorio con Probabilidad Proporcional al "Tamaño" sin Reemplazo.

Las expresiones que determinan la variabilidad de los estimadores, incluyen elementos muy complicados en su cálculo, tales como la probabilidad de que la unidad i de la población esté seleccionada en la muestra (Π_i) y la probabilidad de que dos unidades cualesquiera i y j de la población, estén seleccionadas en la muestra (Π_{ij}), mismos que se encuentran desarrollados por Horvitz y Thompson[2], o bien, por Yates y Grundy[3], dichos elementos, hacen que esta propuesta de muestreo sea muy difícil de implantar, aún para poblaciones relativamente pequeñas como N=20 o N= 30.

Tal parece que la situación antes mencionada no se presenta en la propuesta de Raj, la cuál consiste en:

Considere el muestreo aleatorio con probabilidad al "tamaño" de dos unidades; la primera selección es hecha con probabilidad P_i (i=1, 2, 3, . . . N), las P_i son determinadas con base a las X_i y la segunda selección se efectúa con probabilidad proporcional a los "tamaños" de las unidades restantes; esto es, el elemento j de la población es seleccionado con probabilidad $P_j/(1-P_i)$ si se sabe que previamente el elemento i de la población fue elegido.

Considere los siguientes estimadores del total de la población:

$$t_1 = \frac{y_1}{P_1} \tag{II.13}$$

$$t_2 = y_1 + y_2 \frac{(1-P_1)}{P_2} \tag{II.14}$$

---

[2] Raj, op.cit. pág. 52

[3] Ibid., pág. 54.

donde y_1 y y_2 son los valores de la variable en estudio, asociados con la primera y segunda selección, mientras que P_1 y P_2 son las probabilidades correspondientes a las unidades seleccionadas.

La varianza de t_1 y t_2 estarán dadas por

$$Var(t_1) = \sum_{i=1}^{N}\sum_{j>i}^{N}[X_i X_j (\frac{Y_i}{X_i} - \frac{Y_j}{X_j})^2] \tag{II.15}$$

$$Var(t_2) = \sum_{i=1}^{N}\sum_{j>i}^{N}\left[(1 - \frac{X_i + X_j}{X}) X_i X_j (\frac{Y_i}{X_i} - \frac{Y_j}{X_j})^2\right] \tag{II.16}$$

Se observa que $Var(t_2) < Var(t_1)$; a continuación considérese el estimador de Y, formulado como:

$$t = \frac{t_1 + t_2}{2} \tag{II.17}$$

cuyo valor esperado resulta ser:

$$E[t] = E\left[\frac{t_1 + t_2}{2}\right] = \frac{1}{2}\left(E[t_1] + E[t_2]\right)$$

$$= \frac{1}{2}[Y + Y] = Y \tag{II.18}$$

además

$$E[t_1 t_2] = E_1\left[t_1 E_2\left[t_2 | t_1\right]\right] = Y^2$$

la covarianza entre t_1 y t_2 es cero, por lo tanto:

$$Var(t) = Var\left[\frac{t_1 + t_2}{2}\right] = \frac{1}{4}[Var(t_1) + (Var(t_2))] \qquad \text{(II.19)}$$

como la $Var(t_2) < Var(t_1)$, una cota superior para la $Var(t)$ está dada por:

$$\frac{1}{4}[Var(t_1) + Var(t_1)] = \frac{1}{2}Var(t_1) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j>i}^{N}[X_i X_j (\frac{Y_i}{X_i} - \frac{Y_j}{X_j})^2]$$

esta última expresión, que equivale a la variabilidad que correspondería al estimador del total, en el caso de un muestreo aleatorio con probabilidad proporcional al "tamaño" con reemplazo, para una muestra de dos unidades [4], lo cual implica que:

$$Var(t) = Var[\frac{1}{2}\sum_{i=1}^{2}\frac{y_i}{P_i}] \qquad \text{(II.20)}$$

Lo anterior permite concluir que para n=2, el muestreo aleatorio con probabilidad proporcional al "tamaño" y sin reemplazo, es mejor que el de con reemplazo. Este resultado es generalizado por Des Raj[5], quien considera la selección de una muestra de tamaño n sin reemplazo, en donde la *i*-ésima selección se efectúa con probabilidad proporcional al "tamaño" de las *N*-*i*+1 unidades restantes.

Asimismo, Des Raj propone considerar la secuencia de estimadores siguientes, sea:

$$t_1 = \frac{y_1}{P_1}$$

[4] Raj, op. Cit. fórmula 3.25, pág. 49
[5] Ibid., teorema 3.13 pág. 59.

y en general

$$t_\lambda = y_1 + y_2 + \ldots + y_{\lambda-1} + y_\lambda \frac{1 - \sum_{j=1}^{\lambda-1} P_j}{P_\lambda} \quad \lambda = 2, 3, \ldots, n \tag{II.21}$$

de acuerdo con Des Raj, $E\left[t_\lambda\right] = Y$; que la $E\left[t_\lambda t_\mu\right] = Y^2$ lo cual implica que t_λ y t_μ son no correlacionadas y que $Var(t_\lambda) < Var(t_{\lambda-1})$ para toda λ con base al resultado obtenido en el caso del muestreo de dos unidades.

Como resultado adicional a lo ya contemplado, propone que en virtud de que la $Var(t_n) < Var(t_{n-1}) < \ldots < Var(t_1)$, se considera a

$$t = \frac{t_1 + t_2 + \ldots + t_n}{n} \tag{II.22}$$

como un estimador de Y, y concluye que

$$Var(t) = \frac{1}{n^2} \sum_{i=1}^{n} Var(t_i) < \frac{1}{n} Var(t_1) = Var\Big(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{P_i}\Big) \tag{II.23}$$

determina así que el estimador secuencial, en el caso del muestreo aleatorio con probabilidad proporcional al "tamaño" sin reemplazo, es mejor que el estimador considerado en el caso del mismo muestreo pero con reemplazo y establece un estimador para la $Var(t)$, como sigue:

$$\text{var}(t) = \frac{\sum_{i=1}^{n} (t_i - t)^2}{n(n-1)} \tag{II.24}$$

Con relación a lo antes expuesto, se pueden identificar varios problemas o situaciones de análisis:

i) ¿Sería recomendable considerar a t_n como estimador de Y?, ¿qué tan buena o qué tan mala sería la variabilidad estimada de t_n si esto fuera posible, con relación a la variabilidad de t?, ¿Cuál sería la relación entre las expresiones para las varianzas poblacionales de t y t_i?, para i= 1, 2, 3, . . . , n.

ii) El problema que plantea (i), es que no se conoce ni siquiera una expresión explícita de la $Var(t_n)$, y en consecuencia no se tiene idea de como deba ser un estimador de la misma.

iii) Ciertamente, está demostrado que

$$Var(t) < Var\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}\right)$$

pero ¿cuál es la magnitud que plantea dicha desigualdad?. Generalizando esta última cuestión, ¿qué tanto difieren, de manera práctica, los estimadores considerados en el capítulo anterior para el muestreo aleatorio simple y lo considerado hasta aquí?.

II.4 .- Replanteamiento del Proceso de Estimación que Propone Des Raj.

Si se modifica ligeramente la notación empleada en el proceso de estimación propuesto por Des Raj, de la siguiente manera:

Sea t_j, el estimador para el total poblacional, cuando se realiza la selección j en la misma forma que se propuso en el inciso anterior, considerando además a i_j como el valor de la unidad de la población seleccionada en la j_ésima ocasión y a continuación determinando a:

$$t_1 = \frac{y_{i_1}}{P_{i_1}} \qquad i_1 = \quad 1, 2, 3, \dots, N \tag{II.25}$$

$$t_2 = y_{i_1} + y_{i_2}\frac{1 - P_{i_1}}{P_{i_2}} \qquad i_2 = 1, 2, 3, \dots, N \quad i_2 \neq i_1 \tag{II.26}$$

en general:

$$t_n = \sum_{j=1}^{n-1} y_{i_j} + y_{i_n}\frac{1 - \sum_{j=1}^{n-1} P_{i_j}}{P_{i_n}} \qquad i_n = 1, 2, 3, \dots, N$$

con

$$i_n \neq i_{n-1} \neq i_{n-2} \neq \dots \neq i_1 \tag{II.27}$$

se puede observar que, igual que antes:

$$E\left[t_k\right] = Y \quad \text{para } k = 1, 2, 3, \dots, n \tag{II.28}$$

esto es, t_k es un conjunto secuencial de estimadores insesgados de Y.

Con relación a la variabilidad del conjunto secuencial de estimadores, Des Raj da expresiones explícitas para k=1 y 2, pero no lo hace para cualquier k, cuestión que desarrolló Flores[6] como aquí se presenta

$$Var(t_k) = \sum_{i_1}^{N} P_{i_1} \sum_{i_1 \neq i_2}^{N} \frac{P_{i_2}}{1 - P_{i_1}} \sum_{\substack{i_3 \neq i_1 \\ i_3 \neq i_2}}^{N} \frac{P_{i_3}}{1 - P_{i_1} - P_{i_2}} \dots$$

---

[6] Flores Díaz J. A; Tesis "*Comparación de esquemas muestrales con y sin reemplazo*", 1992, apéndice A

$$\sum_{\substack{i_{k-2} \neq i_1 \\ i_{k-2} \neq i_2 \\ \vdots \\ i_{k-2} \neq i_{k-3}}}^{N} \frac{P_{i_{k-2}}}{1 - \sum_{j=1}^{k-3} P_{i_j}} \sum_{\substack{i_{k-1} \neq i_1 \\ i_{k-1} \neq i_2 \\ \vdots \\ i_{k-1} \neq i_{k-2}}}^{N} \frac{P_{i_{k-1}}}{1 - \sum_{j=1}^{k-2} P_{i_j}}$$

$$\sum_{\substack{i_k \neq i_1 \\ i_k \neq i_2 \\ \vdots \\ i_k \neq i_{k-1}}}^{N} \sum_{\substack{i_k \neq i_1 \\ i_k \neq i_2 \\ \vdots \\ i_k \neq i_{k-1} \\ j_k > i_k}}^{N} X_{i_k} X_{j_k} \left(\frac{Y_{i_k}}{X_{i_k}} - \frac{Y_{j_k}}{X_{j_k}}\right)^2 \tag{II.29}$$

si se considera

$$V_{i_k j_k} = X_{i_k} X_{j_k} \left(\frac{Y_{i_k}}{X_{i_k}} - \frac{Y_{j_k}}{X_{j_k}}\right)^2$$

y se aplica un poco de álgebra se transforma en:

$$Var[t_k] = \sum_{i_k}^{N} \sum_{j_k > i_k}^{N} V_{i_k j_k} [1 - (P_{i_k} + P_{j_k})[1 + \sum_{\substack{i_1 \neq i_k \\ i_1 \neq j_k}}^{N} \frac{P_{i_1}}{1 - P_{i_1}}$$

$$[1 + \sum_{\substack{i_2 \neq i_1 \\ i_2 \neq i_k \\ i_2 \neq j_k}}^{N} \frac{P_{i_2}}{1 - P_{i_1} - P_{i_2}} [1 - \sum_{\substack{i_3 \neq i_1 \\ i_3 \neq i_2 \\ i_3 \neq i_k \\ i_3 \neq j_k}}^{N} \frac{P_{i_3}}{1 - P_{i_1} - P_{i_2} - P_{i_3}}$$

$$[\ldots [1 + \sum_{\substack{i_{k-2} \neq i_1 \\ i_{k-2} \neq i_2 \\ i_{k-2} \neq i_3 \\ \vdots \\ i_{k-2} \neq i_{k-3} \\ i_{k-2} \neq i_k \\ i_{k-2} \neq j_k}}^{N} \frac{P_{i_{k-2}}}{1 - \sum_{j=1}^{k-2} P_{i_j}}] \cdots]]]]] \tag{II.30}$$

Las expresiones anteriores son demasiado complejas y poco fáciles de calcular. Con el propósito de simplificarlas, se propone en (II.30) distribuir sobre todos los elementos dentro de los paréntesis cuadrados, el término externo a ellos, es decir, $V_{i_k j_k}$, desarrollando las sumas y ciertos pasos algebraicos, además de agregar algunas restricciones, de esta manera se tiene:

$$Var[t_k] = \sum_{i_k}^{N} \sum_{j_k > i_k}^{N} V_{i_k j_k} - \sum_{i_k}^{N} \sum_{j_k > i_k}^{N} V_{i_k j_k} (P_{i_k} + P_{j_k}) \sum_{j=1}^{k-2} \left[\prod_{m=1}^{j} \sum_{\substack{i_m \neq i_k \\ i_m \neq j_k \\ i_m \neq i_1 \\ i_m \neq i_2 \\ \vdots \\ i_m \neq i_{m-1}}}^{N} \frac{P_{i_m}}{1 - \sum_{l=1}^{m} P_{i_l}} \right]$$

con

$$\sum_{j=1}^{0} \left[\prod_{m=1}^{j} \sum_{\substack{i_m \neq i_k \\ i_m \neq j_k \\ i_m \neq i_1 \\ i_m \neq i_2 \\ \vdots \\ i_m \neq i_{m-1}}}^{N} \frac{P_{i_m}}{1 - \sum_{l=1}^{m} P_{i_l}} \right] = 1 \quad \text{si} \quad k = 2$$

y

$$\sum_{j=1}^{-1} \left[\prod_{m=1}^{j} \sum_{\substack{i_m \neq i_k \\ i_m \neq j_k \\ i_m \neq i_1 \\ i_m \neq i_2 \\ \vdots \\ i_m \neq i_{m-1}}}^{N} \frac{P_{i_m}}{1 - \sum_{l=1}^{m} P_{i_l}} \right] = 0 \quad \text{si} \quad k = 1 \qquad \text{(II.31)}$$

Aunque relativamente es más sencilla, explícitamente hablando, no es más fácil de evaluar.

Con relación al hecho de que $Var(t_k) < Var(t_{k-1})$, se muestra que:

$$Var[t_k] - Var[t_{k-1}] = -\sum_{i_k}^{N} \sum_{j_k > i_k}^{N} V_{i_k j_k} (P_{i_k} + P_{j_k}) \sum_{\substack{i_1 \neq i_k \\ i_1 \neq j_k}}^{N} \frac{P_{i_1}}{1 - P_{i_1}}$$

$$\sum_{\substack{i_2 \neq i_1 \\ i_2 \neq i_k \\ i_2 \neq j_k}}^{N} \frac{P_{i_2}}{1 - P_{i_1} - P_{i_2}} \cdots \sum_{\substack{i_{k-2} \neq i_1 \\ i_{k-2} \neq i_2 \\ \vdots \\ i_{k-2} \neq i_{k-3} \\ i_{k-2} \neq i_k \\ i_{k-2} \neq j_k}}^{N} \frac{P_{i_{k-2}}}{1 - \sum_{l=1}^{k-2} P_{i_l}} \qquad k = 2, 3, \ldots, n \tag{II.32}$$

la expresión (II.32) es negativa, por tanto, efectivamente, se puede asegurar que

$$Var(t_n) < Var(t_{n-1}) < \ldots < Var(t_1)$$

de la misma manera, es posible demostrar que $Cov[t_i t_j] = 0$ con $i \neq j$ e $i, j = 1, 2, \ldots, n$

En consecuencia, se observa que el estimador $t = \sum_{i=1}^{n} \frac{t_i}{n}$, es insesgado, y que

$$\frac{Var(t_n)}{n} < Var(t) < \frac{Var(t_1)}{n} \tag{II.33}$$

situaciones que responden, en cierto modo, a lo planeado en el punto (i) de las propuestas de análisis.

Igualmente el apartado (ii), ha sido parcialmente resuelto, ya que persiste el problema de obtener un estimador para $Var(t_n)$, y en general para cualquier t_k.

Por último, en relación al punto (iii) se puede decir que sólo es posible lograr una idea sobre las magnitudes con las cuales los estimadores propuestos son, en cada caso, mejores o peores, a través de una simulación en la computadora, tal y como se señaló anteriormente.

Capítulo III

DESCRIPCIÓN DE LA SIMULACIÓN.

III.1.- Introducción.

En los capítulos anteriores, se revisaron los estimadores directo, de razón y de regresión para el caso del muestreo aleatorio simple. En cuanto al muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades, se consideró el estimador correspondiente al total poblacional, cuando el muestreo se desarrolla con reemplazo (el cual por simplicidad será referido como "pptwr") en tanto que, para el muestreo sin reemplazo, se analizó la propuesta de estimación secuencial de Des Raj (referido de ahora en adelante como el estimador de Des Raj o estimador secuencial). En los dos últimos casos, se demostró, que la propuesta de Des Raj es mejor, en términos de una variabilidad menor.

Asimismo, en los resultados planteados al final del primer capítulo, se establece que bajo ciertas condiciones, el estimador directo proporciona mayor variabilidad que el estimador de razón y, a su vez, algo similar sucede entre los estimadores de razón y de regresión.

Aunque existen varios resultados teóricos que establecen ciertas relaciones entre los estimadores considerados, éstos no son suficientes para determinar el comportamiento real de ellos en muy variadas situaciones.

En particular, no se tenía idea de la magnitud en que los resultados antes referidos, se daban bajo diferentes distribuciones de la variable auxiliar y menos todavía se sabía de lo que podría suceder si la relación entre la característica de estudio (Y) y la variable auxiliar era de cierto tipo.

Por consiguiente en la simulación desarrollada por Flores se consideró el muestreo aleatorio simple y el muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades, considerando en cada caso los estimadores ya vistos en los capítulos anteriores y los siguientes puntos:

- Se determinó considerar cuatro patrones funcionales o de relación entre *X* y *Y*, con un número total de 17 casos particulares, que abarcan tanto funciones lineales (6). cuadráticas (3), exponenciales (4) y logarítmicas (4).

- La variable auxiliar *X* se generó, contemplando tres distribuciones: la exponencial. la gausiana y la uniforme.

- Los valores de *X* se desestándarizaron considerando un parámetro de centralidad de 200 y tres niveles de dispersión (2, 10 y 25).

- Con base a lo anterior se determinaron los valores de *Y*.

- Los valores de *N* fueron: 30, 50, 70, 100, 150, 200, 300, 400 y 500.

- Los porcentajes de muestreo fueron: 10, 20, 30, 40 y 45 por ciento.

Otros detalles del estudio pueden verse directamente en la tesis "Comparaciones de esquemas muestrales con y sin reemplazo", sin embargo lo anterior plantea las características primordiales de la simulación cuyos resultados a "grosso modo" fueron: que es muy recomendable utilizar información auxiliar si es que es posible contar con ella y en dicho caso implantar el estimador de la media con el enfoque de regresión, lo anterior es recomendable aún en el caso de que la otra alternativa fuera implantar un muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades.

Con relación a la suposición de normalidad ésta fue mejor observada en el caso del estimador de razón en el muestreo aleatorio simple y fue más mala en el estimador de regresión.

Para el muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades el estimador con reemplazo fue más normal que el secuencial.

Lo que fue la base para el desarrollo del presente trabajo, consistió en observar que las características (X, Y), generadas según se explicó líneas arriba, tuvieron altos grados de correlación positiva que oscilaron entre 99.997 y 100 ó (–100) por ciento.

Por consiguiente la simulación del presente trabajo consiste en:

Los tamaños poblacionales oscilaron entre 30 y 350, los porcentajes de muestreo implantados fueron del 10 al 45 por ciento de la población.

A fin de comparar las diferentes variabilidades se efectúo una simulación, donde para una cierta población N, se conocen los valores de la variable auxiliar X, mientras que los valores de la característica en estudio Y, son desconocidos, sin embargo X y Y están relacionados con una correlación específica y bajo estas condiciones se realizó el cálculo de los estimadores ya presentados y las estimaciones de sus varianzas correspondientes.

III.2.- Los Datos de Entrada para la Simulación.

Para la variable auxiliar o independiente X se generaron N valores aleatorios desestándarizados, a partir de los cuales se obtuvieron las variables dependientes con una covarianza de 0.3, 0.5 y 0.8.

Los tamaños poblacionales o valores considerados de N fueron 30, 50, 70, 100, 200 y 350, siendo los porcentajes de muestreo 10, 20, 30, 40 y 45. Cabe señalar que cuando el número de elementos de la muestra resultó ser fraccionario, éste se redondeo al entero mayor.

III.3.- Los Datos de Salida de la Simulación.

A partir de la información producida por la simulación, se calcularon la media y varianza muestral de cada uno de los cinco procedimientos de estimación considerados, también se generaron tablas con los valores poblacionales (media y varianza) de cada modelo para todas las combinaciones de elementos contemplados en el punto anterior.

De igual forma, se crearon, para todos los elementos, archivos con valores de la variable independiente X y la dependiente Y.

III.4.- Programas.

Se utilizaron dos programas para llevar a cabo la simulación, uno de ellos, simuló el muestreo aleatorio simple (NOPPT) y, el otro, el muestreo con probabilidad proporcional al "tamaño" de las unidades (PPT).

Las subrutinas y demás elementos que los componen se encuentran explicados en los comentarios agregados a cada uno de ellos. Ver apéndices C.1 y C.2.

Los programas fueron desarrollados en FORTRAN77 y se emplearon algunas rutinas del paquete IMSL[1], tales como el generador de las X_i (RNSET) para la generación de las Y_i se utilizaron las rutinas (UMACH, CHFAC, RNMVN) por lo que se refiere al proceso de selección en el caso del muestreo con probabilidades iguales, se emplearon las rutinas (RNSET, RNSRI y RNGET); Mientras que para el programa de muestreo con probabilidades proporcionales al "tamaño", se utilizaron las rutinas (RNUNF, GGUBFS y RNGET).

Además de los programas ya referidos que simularon los procesos de muestreo y estimación, también se desarrollaron programas en (SPSSX) que sirvieron para obtener

[1] International Mathematical System Library

resultados estadísticos descriptivos, como la prueba de Kolmogorov-Smirnov para normalidad, los cuales fueron aplicados a los resultados muestrales obtenidos para cada modelo y en cada tipo de muestreo. Ver apéndice C.3.

RESULTADOS Y COMENTARIOS PARCIALES.

IV.1.- Introducción.

A continuación se referirán los datos principales de la simulación efectuada y comentarios de las tablas presentadas en el apéndice B.

IV.2.-Ejecución de la Simulación.

Los programas que simulan cada uno de los muestreos considerados, se corrieron en CRAY, generando un archivo de resultados muestrales para cada modelo de relación, entre la variable auxiliar X y la variable en estudio Y, siempre y cuando la variabilidad poblacional no fuera cero, el número máximo de registros elaborados para un modelo fue de 11000 para los archivos creados por el programa (NOPPT) que realizó la simulación del muestreo aleatorio simple, constando éstos de 122 caracteres aproximadamente, el número de registros para el programa (PPT) que realizó la simulación de los modelos para el muestreo con probabilidad proporcional es de 8000 con 100 caracteres aproximadamente, donde cada uno de los registros creados por dichos programas contenía la información correspondiente al número de muestra, porcentaje de muestreo, valores de la media estimada, la estimación de la variabilidad del estimador en cuestión y su covarianza respectiva para los procesos de estudio dando un total de 57000 registros generados.

IV.3.- Corridas con el SPSSX.

Para el análisis estadístico de la información generada por los dos programas que simularon los tipos de muestreo referidos, se realizaron cuatro programas.

El primero de ellos se desarrolló para obtener resultados de las medias muéstrales. Para ello se utilizó la rutina que obtiene frecuencias, generando estadísticas de las poblaciones analizadas, en este caso las media, varianza, la moda, la mediana, el valor máximo y el valor mínimo.

El segundo programa desarrollado, obtuvo resultados semejantes para las varianzas muestrales con el uso de la misma rutina.

El tercero se empleó para generar los archivos con las medias muéstrales estándarizadas, y poder aplicar, posteriormente, a dichos datos el último programa desarrollado.

El cuarto programa utilizó la rutina de pruebas estadísticas no parámetricas (npar-tes) particularmente el caso de la prueba de Kolmogorov-Smirnov, para ver lo relacionado a la suposición de normalidad que a menudo se considera en el ejercicio del muestreo.

IV.4.-Los Resultados Obtenidos de las Corridas con el SPSSX.

Para el primer modelo que tiene las variables "*X*" y "*Y*" con una covarianza de 0.3, se observó que la media estimada más cercana a la media poblacional, para las poblaciones de 70 y 100 fue la media estimada por el estimador de regresión, para las poblaciones de 50 y 350 la media más cercana fue la media estimada por el estimador simple, para la población de 200 el valor más cercano fue la media estimada por el estimador de razón y para la población de 30 el mejor estimador fue el secuencial. Lo que se observó con respecto a la varianza, fue que la varianza estimada por el método de regresión fue la más pequeña con respecto a los otros estimadores o métodos, la varianza más alta fue la obtenida del estimador secuencial, que tiene un incremento del 100% aproximadamente con respecto a la varianza obtenida por el estimador de regresión, esto se presentó en cada uno de los diferentes tamaños de población.

Para el segundo modelo, el de covarianza 0.5, se observó que el valor más cercano a la media poblacional en las poblaciones de 50, 70, 100, 200, fue la media estimada por el estimador de regresión, en las poblaciones de 30 y 350 la media estimada más cercana a la media poblaciones fue la que se obtuvo a través del estimador simple. Con respecto a la varianza, de la misma forma que en el modelo anterior, el estimador que presentó menor variabilidad fue el de regresión y el de mayor varianza fue el secuencial, que tuvo un incremento mayor al 50% con respecto a la varianza estimada por el estimador de regresión, este incremento fue creciendo al aumentar la fracción de muestreo llegando a ser mayor de 100%.

Para el tercer modelo, con covarianza de 0.8, se observó que el estimador más cercano a la media poblacional en las poblaciones de 30, 100 fue el de la media estimada por el estimador de razón, la media estimada por el estimador de regresión fue la más cercana a la media poblacional para las poblaciones de 50 y 70, para la población de 200 el mejor estimador fue el de la media estimada por el secuencial y para la población de 350 el mejor estimador fue el que se obtuvo a través del estimador de pptwr. Este modelo se comportó en forma distinta a los otros modelos con respecto a la varianza, ya que el estimador simple, fue el que tuvo una mayor varianza, el que presentó una menor variabilidad fue el estimador de regresión, el incremento que se presentó entre los dos estimadores con respecto a la varianza fue del 150% al 200% aproximadamente, ésta fue creciendo conforme aumentó la fracción de muestreo.

Las Estadísticas calculadas para la prueba de Kolmogorov-Smirnov.

En las tablas correspondientes a las estadísticas calculadas para la prueba de Kolmogorov-Smirnov por modelo, se puede observar que en todos los modelos se acepta la hipótesis de que la muestra aleatoria tiene una distribución normal, con un nivel de significancia de $\alpha=0.05$. Para un nivel de significancia de $\alpha = 0.1$, sólo en dos casos se rechazó la hipótesis de normalidad: para el modelo de Cov.=0.3 con N = 350 y fracción de muestreo de 0.40 y en el modelo de Cov. = 0.8 con N = 350 y fracción de muestreo de 0.45 ambos casos en el estimador secuencial. Al aumentar el nivel de significancia a $\alpha = 0.5$, en los modelos de Cov. = 0.3 y Cov.= 0.5 sólo se rechazó la hipótesis de normalidad en el 44% de los casos para N = 350 contemplando todos los estimadores, para el modelo Cov. = 0.8 se rechazó la prueba de hipótesis para N = 200 en un 20% para los estimadores pptwr, secuencial y regresión, para el estimador simple se rechazó en un 40% y en el estimador de razón no se rechazó la hipótesis de normalidad. Con N = 350 en todos los estimadores se presentaron casos en la que se rechazó la hipótesis de normalidad, para los estimadores pptwr y secuencial se rechazó la hipótesis en un 80% de los casos estudiados , para el estimador de regresión y razón se rechazó el 60% de los casos y para el estimador simple sólo se rechazó en un 40% de los casos. Con esto se puede concluir que se acepta la hipótesis de normalidad para los modelos en estudio.

Con relación al cuadro-resumen por modelo, que se presenta, algunas de las consideraciones que se pueden hacer son las siguientes:

Muestreo aleatorio simple:

I. El estimador de razón presentó un porcentaje mayor en número de veces, el valor más pequeño para el modelo de Cov = 0.3 en las estadísticas de Kolmogorov-Smirnov.

II. El estimador simple resultó tener un porcentaje mayor en número de veces, el valor más pequeño de las estadísticas de Kolmogorov-Smirnov para el modelo de Cov = 0.5.

III. El estimador de regresión resultó tener el mismo porcentaje de veces, el valor más pequeño que el estimador simple para el modelo Cov = 0.8.

IV. Para el caso del valor más grande de la estadística, el porcentaje de número de veces del estimador directo, razón y regresión fue el mismo para los modelos Cov = 0.3 y Cov = 0.8.

El estimador de regresión resultó tener el porcentaje mayor en número de veces, el valor más pequeño de la estadística.

El muestreo aleatorio con probabilidades proporcionales al "tamaño" de las unidades presentó el siguiente comportamiento:

1. El estimador pptwr, resultó ser el que mostró mayor número de veces, el valor más pequeño de la estadística en el modelo Cov = 0.3.

2. El estimador pptwr y secuencial presentaron el mismo número de veces, el valor más pequeño de la estadística en los modelos Cov = 0.5 y Cov = 0.8.

3. El estadístico más grande se presentó en igual porcentaje para el estimador pptwr y secuencial para los modelos Cov = 0.3 y Cov = 0.8.

4. En el modelo Cov = 0.5 el estimador secuencial presentó el porcentaje mayor en número de veces el valor más grande del estadístico.

Capítulo V

CONCLUSIONES Y RECOMENDACIONES.

En virtud del trabajo hasta aquí planteado, se puede resumir todo lo anterior en dos sentidos: uno teórico y otro práctico.

En lo teórico se considera que el replanteamiento del estimador propuesto por Des Raj o secuencial, representa un avance ligero en el problema en cuestión toda vez, que aunque ya se logró una o varias expresiones para la variabilidad de t_j y en particular de t_n, éstas continúan representando enormes dificultades de implantación sobre todo cuando n y N crecen, y el problema de estimar la variabilidad, para cualquier j, pero principalmente para $j = n$ está desarrollado en la tesis "Comparaciones de esquemas muestrales con y sin reemplazo" desarrollado por el M. en C. José Antonio Díaz Flores.

En el aspecto práctico uno de los resultados que se obtuvieron en la simulación es que, resulta ser más importante considerar la información adicional de manera que se contemple un muestreo aleatorio simple y se calcule un estimador de regresión, a utilizar dicha información para muestrear con probabilidad proporcional al "tamaño" de las unidades.

Un segundo punto es que si se tiene información adicional de una población y se va a muestrear con probabilidad proporcional al "tamaño" de las unidades, el hacerlo con reemplazo o sin reemplazo no proporciona una ventaja con relación a la variabilidad del estimador de regresión en el muestreo con probabilidades iguales, ya que su varianza es mayor en un 100% aproximadamente al de la varianza del estimador de regresión.

Por lo que corresponde a la suposición de normalidad en los estimadores de la media, la simulación realizada señala que para el caso del muestreo aleatorio simple con probabilidad igual, el estimador de regresión es el que mejor se comporta con la hipótesis en cuestión, luego seguiría el directo y por último el de razón.

Para el muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades. puede considerarse que el estimador pptwr es el que mejor se comporta con la hipótesis de normalidad y le seguiría el secuencial.

Por lo que se puede concluir que el mejor estimador es el de regresión, por presentar una menor variabilidad y tener el mejor comportamiento con respecto a la hipótesis de normalidad y que no importa el grado de correlación que exista entre las variables X y Y, ya que en todos los casos estudiados presentan el mismo comportamiento, con respecto a los modelos considerados sería recomendable estudiar modelos drásticamente diferentes.

APÉNDICE A

En este apéndice se demostraron algunos resultados utilizados en los capítulos anteriores

Demostración de $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \bar{Y})^2 - n(\bar{y} - \bar{Y})^2$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left[(y_i - \bar{Y}) - (\bar{y} - \bar{Y})\right]^2$$

$$= \sum_{i=1}^{n}\left[(y_i - \bar{Y})^2 - 2(y_i - \bar{Y})(\bar{y} - \bar{Y}) + (\bar{y} - \bar{Y})^2\right]$$

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

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

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

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

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

por lo tanto

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \bar{Y})^2 - n(\bar{y} - \bar{Y})^2 \qquad \text{(A.1)}$$

Demostración de $E\left[s_y^2\right] = S_y^2$

$$E\left[s_y^2\right] = E\left[\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}\right]$$ aplicando el resultado anterior (A.1)

$$= E\left[\frac{\sum_{i=1}^{n} (y_i - \bar{Y})^2 - n(\bar{y} - \bar{Y})^2}{n-1}\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2 - n(\bar{y} - \bar{Y})^2\right]$$

$$= \frac{1}{n-1}\left[E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2\right] - E\left[n(\bar{y} - \bar{Y})^2\right]\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E\left[(y_i - \bar{Y})^2\right] - nE\left[(\bar{y} - \bar{Y})^2\right]\right]$$

$$= \frac{1}{n-1}\left[nE\left[(y_i - \bar{Y})^2\right] - nE\left[(\bar{y} - \bar{Y})^2\right]\right]$$

$$= \frac{1}{n-1}\left[nVar(y_i) - nVar(\bar{y})\right]$$

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

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

$$= \frac{n\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N(n-1)}\left[\frac{Nn - n - N + n}{n(N-1)}\right]$$

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

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

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

por lo tanto

$$E\left[s_y^2\right] = S_y^2 \qquad \text{(A.2)}$$

Demostración de $E\left[s_{xy}\right] = S_{xy}$

$$E\left[s_{xy}\right] = \frac{1}{n-1} E\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n}\left[(x_i - \bar{x} + \bar{X} - \bar{X})(y_i - \bar{y} + \bar{Y} - \bar{Y})\right]\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n}\left[((x_i - \bar{X}) - (\bar{x} - \bar{X}))((y_i - \bar{Y}) - (\bar{y} - \bar{Y}))\right]\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n}\left[(x_i - \bar{X})(y_i - \bar{Y}) - (\bar{x} - \bar{X})(y_i - \bar{Y}) - (x_i - \bar{X})(\bar{y} - \bar{Y}) + (\bar{x} - \bar{X})(\bar{y} - \bar{Y})\right]\right]$$

$$=\frac{1}{n-1}\left[nCov(x,y)+nCov(\bar{x},\bar{y})-E\left[\sum_{i=1}^{n}\left[(\bar{x}-\bar{X})(y_i-\bar{Y})+(x_i-\bar{X})(\bar{y}-\bar{Y})\right]\right]\right]$$

$$=\frac{1}{n-1}\left[nCov(x,y)+nCov(\bar{x},\bar{y})-E\left[(\bar{x}-\bar{X})(n\bar{y}-\bar{Y})+(n\bar{x}-\bar{X})(\bar{y}-\bar{Y})\right]\right]$$

$$=\frac{1}{n-1}\left[nCov(x,y)+nCov(\bar{x},\bar{y})-E\left[2\bar{X}\bar{Y}+2n\bar{x}\bar{y}-(n+1)\bar{x}\bar{Y}-(n+1)\bar{y}\bar{X}\right]\right]$$

$$=\frac{1}{n-1}\left[nCov(x,y)+nCov(\bar{x},\bar{y})-2\bar{X}\bar{Y}-2n\left(\bar{X}\bar{Y}+Cov(\bar{x},\bar{y})\right)+(n+1)\bar{X}\bar{Y}+(n+1)\bar{X}\bar{Y}\right]$$

$$=\frac{1}{n-1}\left[nCov(x,y)+nCov(\bar{x},\bar{y})-2\bar{X}\bar{Y}-2n\bar{X}\bar{Y}-2nCov(\bar{x},\bar{y})+(n+1)\bar{X}\bar{Y}+(n+1)\bar{X}\bar{Y}\right]$$

$$=\frac{1}{n-1}\left[nCov(x,y)-nCov(\bar{x},\bar{y})-2(n+1)\bar{X}\bar{Y}+2(n+1)\bar{X}\bar{Y}\right]$$

$$=\frac{n}{n-1}\left[\frac{\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{N}-\frac{(N-n)\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{nN(N-1)}\right]$$

$$=\frac{n\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{N(n-1)}\left[1-\frac{(N-n)}{n(N-1)}\right]$$

$$= \frac{n\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{N(n-1)}\left[\frac{N(n-1)}{n(N-1)}\right]$$

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

por lo tanto se puede concluir

$$E\left[s_{xy}\right] = S_{xy} \qquad \text{(A.3)}$$

Demostración de $Cov(\hat{R}, \overline{X}) = 0$

$$Cov(\hat{R}, \overline{X}) = E\left[\left(\hat{R} - E\left[\hat{R}\right]\right)\left(\overline{X} - E\left[\overline{X}\right]\right)\right]$$

$$= E\left[\left(\hat{R} - E\left[\hat{R}\right]\right)\left(\overline{X} - \overline{X}\right)\right]$$

$$= E\left[\left(\hat{R} - E\left[\hat{R}\right]\right)0\right]$$

$$= 0$$

por lo tanto

$$Cov(\hat{R}, \overline{X}) = 0 \qquad \text{(A.4)}$$

Demostración de $E\left[\hat{\bar{Y}}_{Ra}\right] = \bar{X}E\left[\hat{R}\right]$

$$E\left[\hat{\bar{Y}}_{Ra}\right] = E\left[\hat{R}\bar{X}\right]$$

$$= E\left[\hat{R}\right]E\left[\bar{X}\right] - Cov(\hat{R}, \bar{X})$$

$$= \bar{X}E\left[\hat{R}\right] - Cov(\hat{R}, \bar{X}) \quad \text{aplicando el resultado (A.4)}$$

$$= \bar{X}E\left[\hat{R}\right]$$

por lo tanto

$$E\left[\hat{\bar{Y}}_{Ra}\right] = \bar{X}E\left[\hat{R}\right] \tag{A.5}$$

Demostración de $E\left[\hat{R}\right] \cong R$

$E\left[\hat{R}\right] = E\left[\dfrac{\bar{y}}{\bar{x}}\right]$ para obtener la esperanza de este cociente se considera:

$$f(\theta)=E\left[\frac{\bar{y}}{\bar{X}+\theta(\bar{x}-\bar{X})}\right] \tag{A.6}$$

se observa que en particular para $\theta = 1$ se obtiene $f(1)=E\left[\frac{\bar{y}}{\bar{x}}\right]$

Utilizando el teorema de Taylor que consiste en:

$f(b)=f(a)+(b-a)f'(a)+\frac{(b-a)^2}{2!}f''(a)+\ldots+\frac{(b-a)^n}{n!}f^n(a)+k$ donde k es el residuo.

Se obtienen las derivadas de $f(\theta)$

$f'(\theta)=E\left[\frac{-\bar{y}(\bar{x}-\bar{X})}{\left[\bar{X}-\theta(\bar{x}-\bar{X})\right]^2}\right]$ primera derivada de $f(\theta)$

$f''(\theta)=E\left[\frac{2\bar{y}(\bar{x}-\bar{X})^2}{\left[\bar{X}-\theta(\bar{x}-\bar{X})\right]^3}\right]$ segunda derivada de $f(\theta)$

$f'''(\theta)=E\left[\frac{-6\bar{y}(\bar{x}-\bar{X})^3}{\left[\bar{X}-\theta(\bar{x}-\bar{X})\right]^4}\right]$ tercera derivada de $f(\theta)$

$f^i(\theta)=E\left[\frac{(-1)^i i!\,\bar{y}(\bar{x}-\bar{X})^i}{\left[\bar{X}-\theta(\bar{x}-\bar{X})\right]^{i+1}}\right]$ i_esíma derivada de $f(\theta)$

por consiguiente para $a=\theta=0$ y $b=1$

$$f(1)=f(0)+\left(-E\left[\frac{-\bar{y}(\bar{x}-\bar{X})}{\left[\bar{X}+\theta(\bar{x}-\bar{X})\right]^2}\right]\right)+$$

$$\left(-E\left[\frac{2\bar{y}(\bar{x}-\bar{X})^2}{2\left[\bar{X}+\theta(\bar{x}-\bar{X})\right]^3}\right]\right)+\ldots+\left(-E\left[\frac{(-1)^i i!\,\bar{y}(\bar{x}-\bar{X})^i}{i!\left[\bar{X}+\theta(\bar{x}-\bar{X})\right]^{i+1}}\right]\right)$$

$f(1)=f(0)-\sum_{i=1}^{\infty}E\left[\frac{(-1)^i i!\,\bar{y}(\bar{x}-\bar{X})^i}{i!\left[\bar{X}+\theta(\bar{x}-\bar{X})\right]^{i+1}}\right]$ que es equivalente a la siguiente expresión

$$f(1)\cong E\left[\frac{\bar{y}}{\bar{X}}\right]-\sum_{i=1}^{\infty}E\left[\frac{(-1)^i\,\bar{y}(\bar{x}-\bar{X})^i}{\bar{X}^{i+1}}\right] \tag{A.7}$$

en la cual suele considerarse como un valor insignificante el segundo término de la derecha de la expresión (A.7) con esto se obtiene la siguiente expresión:

$$f(1)\cong\frac{E[\bar{y}]}{\bar{X}}=\frac{\bar{Y}}{\bar{X}} \tag{A.8}$$

por lo tanto se concluye que

$$E\left[\hat{R}\right]\cong\frac{\bar{Y}}{\bar{X}}=R \tag{A.9}$$

Obtención de la esperanza de $E\left[\left(\frac{\bar{y}-R\bar{x}}{\bar{x}}\right)^2\right]$

Por ser $E\left[\left(\frac{\bar{y}-R\bar{x}}{\bar{x}}\right)^2\right]$ la esperanza de un cociente se considera a

$f(\theta)=E\left[\frac{(\bar{y}-R\bar{x})^2}{\bar{X}^2-\theta(\bar{x}^2-\bar{X}^2)}\right]$ que en particular para $f(1)$ se obtiene $E\left[\left(\frac{\bar{y}-Rx}{\bar{x}}\right)^2\right]$

Se obtienen las derivadas de $f(\theta)$

$f'(\theta)=E\left[\frac{-(\bar{y}-R\bar{x})^2(\bar{x}^2-\bar{X}^2)}{\left[\bar{X}^2-\theta(\bar{x}^2-\bar{X}^2)\right]^2}\right]$ primera derivada de $f(\theta)$

$f''(\theta)=E\left[\frac{2(\bar{y}-R\bar{x})^2(\bar{x}^2-\bar{X}^2)^2}{\left[\bar{X}^2-\theta(\bar{x}^2-\bar{X}^2)\right]^3}\right]$ segunda derivada de $f(\theta)$

$f'''(\theta)=E\left[\frac{-6(\bar{y}-R\bar{x})^2(\bar{x}^2-\bar{X}^2)^3}{\left[\bar{X}^2-\theta(\bar{x}^2-\bar{X}^2)\right]^4}\right]$ tercera derivada de $f(\theta)$

$f^{i}(\theta)=E\left[\frac{(-1)^i i!(\bar{y}-R\bar{x})^2(\bar{x}^2-\bar{X}^2)^i}{\left[\bar{X}^2-\theta(\bar{x}^2-\bar{X}^2)\right]^{i+1}}\right]$ i_esíma derivada de $f(\theta)$

por consiguiente para $a=\theta=0$ y $b=1$ aplicando el Teorema de Taylor se obtiene

$$f(1)=E\left[\frac{(\bar{y}-R\bar{x})^2}{\bar{X}^2}\right]-\sum_{i=1}^{\infty}E\left[\frac{(-1)^i(\bar{y}-R\bar{x})^2(\bar{x}^2-\bar{X}^2)^i}{(\bar{X}^2)^{i+1}}\right] \tag{A.10}$$

se considera como un valor insignificante el segundo término de la derecha de la expresión (A.10) con esto se obtiene el siguiente resultado:

$$E\left[\frac{(\bar{y}-R\bar{x})^2}{\bar{x}^2}\right]\cong E\left[\frac{(\bar{y}-R\bar{x})^2}{\bar{X}^2}\right]$$

$$=\frac{E\left[(\bar{y}-R\bar{x})^2\right]}{\bar{X}^2} \tag{A.11}$$

Estimación de $ECM(\hat{\bar{Y}})=E\left[(\hat{R}\bar{X}-\bar{Y})^2\right.$

$$E\left[(\hat{R}\bar{X}-\bar{Y})^2\right]=E\left[\left(\frac{\bar{y}}{\bar{x}}\bar{X}-\bar{Y}\right)^2\right]$$

$$=E\left[\bar{X}^2\left(\frac{\bar{y}-R\bar{x}}{\bar{x}}\right)^2\right]$$

$$=\bar{X}^2E\left[\left(\frac{\bar{y}-R\bar{x}}{\bar{x}}\right)^2\right]$$ utilizando el resultado (A.11)

$$\cong \bar{X}^2 \frac{E\left[(\bar{y} - R\bar{x})^2\right]}{\bar{X}^2}$$

$$= E\left[(\bar{y} - R\bar{x})^2\right]$$

por lo tanto se puede decir que

$$E\left[(\hat{R}\bar{X} - \bar{Y})^2\right] \cong E\left[(\bar{y} - R\bar{x})^2\right] \qquad \text{(A.12)}$$

Demostración de $E\left[(\bar{y} - R\bar{x})^2\right] = \frac{(1 - \frac{n}{N})}{n}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right]$

Sea $z_i = y_i - Rx_i$ que tiene las siguientes características $\bar{Z} = \sum_{i=1}^{N} z_i = \bar{Y} - R\bar{X} = 0$,

$\bar{z} = \sum_{i=1}^{n} z_i = \bar{y} - R\bar{x}$, $E[\bar{z}] = \bar{Z} = 0$ y $S_Z^2 = \sum_{i=1}^{N} \frac{(z_i - \bar{Z})^2}{N - 1}$

Se calcula la $Var(\bar{z})$

$Var(\bar{z}) = E\left[(\bar{z} - E[\bar{z}])^2\right]$ aplicando el teorema[1] 2.2 se obtiene

[1] Cochcran , *Técnicas de Muestreo*, 1992, ver capítulo 2, teorema 2.2, pág. 47.

$$= (1 - \frac{n}{N})\frac{S_Z^2}{n} \qquad \text{(A.13)}$$

con $S_Z^2 = \sum_{i=1}^{N} \frac{(z_i - \bar{Z})^2}{N-1}$ pero como una de sus características es $\bar{Z} = 0$, entonces

$$S_Z^2 = \sum_{i=1}^{N} \frac{(z_i)^2}{N-1}$$

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

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}\left(y_i^2 - 2Rx_i y_i + R^2 x_i^2\right)\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - 2R\sum_{i=1}^{N} x_i y_i + R^2 \sum_{i=1}^{N} x_i^2\right]$$

agregando $-N\bar{Y}^2 + 2RN\bar{Y}\bar{X} - R^2 N\bar{X}^2$ lo cual vale cero

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\bar{Y}^2 - 2R(\sum_{i=1}^{N} x_i y_i - N\bar{Y}\bar{X}) + R^2(\sum_{i=1}^{N} x_i^2 - N\bar{X}^2)\right]$$

$$= \frac{1}{N-1}\left[(N-1)S_y^2 - 2(N-1)RS_{xy} + R^2(N-1)S_x^2\right]$$

por lo tanto

$$S_z^2 = S_y^2 - 2RS_{xy} + R^2 S_x^2 \tag{A.14}$$

volviendo al cálculo de la $Var(\bar{z})$ se retoma la expresión (A.13) y se sustituye el resultado de la expresión (A.14)

$$Var(\bar{z}) = \frac{(1 - \frac{n}{N})}{n}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right]$$

como la $Var(\bar{z}) = E\left[(\bar{y} - R\bar{x})^2\right]$ por las características de z_i entonces

$$E\left[(\bar{y} - R\bar{x})^2\right] = \frac{(1 - \frac{n}{N})}{n}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right] \tag{A.15}$$

retomando los resultados (A.12), (A.15) se concluye lo siguiente:

$$E\left[(\hat{R}\bar{X} - \bar{Y})^2\right] \cong \frac{(1 - \frac{n}{N})}{n}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right] \tag{A.16}$$

Obtención de la $E\left[\hat{\bar{Y}}_{Re}\right]$

$$E\left[\hat{\bar{Y}}_{Re}\right] = E\left[\bar{y} + b(\bar{X} - \bar{x})\right]$$

$$= E\left[\bar{y}\right] + E\left[b(\bar{X} - \bar{x})\right]$$

$$= \bar{Y} + E\left[b(\bar{X} - \bar{x}) - B(\bar{X} - \bar{x})\right]$$

$$= \bar{Y} + E\left[(b - B)(\bar{X} - \bar{x})\right]$$

$$= \bar{Y} - E\left[(b - B)(\bar{x} - \bar{X})\right]$$

$$= \bar{Y} - Cov(b, \bar{x}) \qquad \text{(A.17)}$$

Para obtener el sesgo de la estimación de regresión lineal, se introduce la variable auxiliar e_i definida por la relación

$$e_i = y_i - \bar{Y} - B(x_i - \bar{X})$$

con las propiedades de $\sum_{i=1}^{N} e_i = 0$ y $\sum_{i=1}^{N} e_i(x_i - \bar{X}) = 0$, y por definición de B, ahora

$$b = \sum_{i=1}^{n} y_i(x_i - \bar{x}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} \left[\bar{Y} + B(x_i - \bar{X}) + e_i\right](x_i - \bar{x}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= B + \sum_{i=1}^{n} e_i(x_i - \bar{x}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{A.18}$$

se reemplaza $\sum_{i=1}^{n} (x_i - \bar{x})^2 = (n-1)s_x^2$ en la expresión (A.18) y se escribe

$$\sum_{i=1}^{n} e_i(x_i - \bar{x}) = \sum_{i=1}^{n} e_i(x_i - \bar{X}) + n\bar{e}(\bar{X} - \bar{x}) \tag{A.19}$$

sustituyendo las expresiones (A.18) y (A.19) en la siguiente expresión se obtiene:

$$E\left[b(\bar{x} - \bar{X})\right] = E\left[\frac{\sum_{i=1}^{n} e_i(x_i - \bar{X})(\bar{X} - \bar{x}) + n\bar{e}(\bar{X} - \bar{x})^2}{(n-1)s_x^2}\right]$$

$$= E\left[\frac{\sum_{i=1}^{n} e_i (x_i - \overline{X})(\overline{X} - \bar{x})}{(n-1)s_x^2}\right] + E\left[\frac{n\bar{e}(\overline{X} - \bar{x})^2}{(n-1)s_x^2}\right]$$ por las

características de e_i

$$= E\left[\frac{\sum_{i=1}^{n} e_i (x_i - \overline{X})(\overline{X} - \bar{x})}{(n-1)s_x^2}\right] \tag{A.20}$$

para obtener la esperanza de la expresión (A.20) por ser un cociente se utiliza el teorema de Taylor, entonces se define a $f(\theta)$

$$f(\theta) = E\left[\frac{\sum_{i=1}^{n} e_i (x_i - \overline{X})(\overline{X} - \bar{x})}{(n-1)S_x^2 + \theta\left((n-1)s_x^2 - (n-1)S_x^2\right)}\right]$$

su i_ésima derivada esta determinada por la siguiente expresión:

$$f^i(\theta) = E\left[\frac{(-1)^i i! \sum_{i=1}^{n} e_i (x_i - \overline{X})(\bar{x} - \overline{X})}{\left[(n-1)S_x^2 + \theta\left((n-1)s_x^2 - (n-1)S_x^2\right)\right]^{i+1}}\right] \tag{A.21}$$

por lo tanto una aproximación del orden cero para la esperanza de la expresión (A.20) es

$$E\left[\frac{\sum_{i=1}^{n} e_i(x_i - \overline{X})(\overline{X} - \bar{x})}{(n-1)s_x^2}\right] = \frac{1}{(n-1)S_x^2}\left[E\left[\sum_{i=1}^{n} e_i(x_i - \overline{X})(\overline{X} - \bar{x})\right] + \sum_{i=1}^{\infty} E\left[(-1)^i \sum_{i=1}^{n} e_i(x_i - \overline{X})(\bar{x} - \overline{X})\right]\right]$$

que es

$$E\left[\frac{\sum_{i=1}^{n} e_i(x_i - \overline{X})(\overline{X} - \bar{x})}{(n-1)s_x^2}\right] \cong 0$$

por lo tanto se puede concluir que:

$$E\left[b\left(\bar{x} - \overline{X}\right)\right] \cong 0 \tag{A.22}$$

en consecuencia

$$E\left[\hat{\overline{Y}}_{\text{Re}}\right] \cong \overline{Y} \tag{A.23}$$

Obtención de la $E[b]$

$$E[b] = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] \tag{A.24}$$

$$= E\left[\frac{\sum_{i=1}^{n}\left(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}\right)}{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}\right]$$

$$= E\left[\frac{\left(\sum_{i=1}^{n}x_i y_i - \bar{y}\sum_{i=1}^{n}x_i - \bar{x}\sum_{i=1}^{n}y_i + n\bar{x}\bar{y}\right)}{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}\right]$$

$$= E\left[\frac{\sum_{i=1}^{n}x_i y_i - n\bar{x}}{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}\right]$$ para obtener la esperanza de este cociente se considera a:

$$f(\theta) = E\left[\frac{\sum_{i=1}^{n}x_i y_i - n\bar{x}}{\left(\sum_{i=1}^{N}x_i^2 - N\bar{X}^2\right) - \theta\left\{\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right) - \left(\sum_{i=1}^{N}x_i^2 - N\bar{X}^2\right)\right\}}\right]$$ que en

particular para $\theta = 1$ se obtiene $f(1) = E\left[\dfrac{\sum_{i=1}^{n} x_i y_i - nx}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\right]$ nuevamente se utiliza el Teorema de Taylor

Se obtienen las respectivas derivadas de $f(\theta)$

$$f'(\theta) = E\left[\frac{-\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\right)}{\left[\left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right) - \theta\left\{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) - \left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right)\right\}\right]^2}\right] \text{ primera derivada}$$

$$f''(\theta) = E\left[\frac{2\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)^2}{\left[\left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right) - \theta\left\{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) - \left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right)\right\}\right]^3}\right] \text{ segunda derivada}$$

$$f'''(\theta) = E\left[\frac{-6\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)^3}{\left[\left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right) - \theta\left\{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) - \left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right)\right\}\right]^4}\right] \text{ tercera derivada}$$

$$f^{i}(\theta)=E\left[\frac{(-1)^{i}i!\left(\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right)^{i}}{\left[\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)-\theta\left\{\left(\sum_{i=1}^{n}x_{i}^{2}-n\bar{x}^{2}\right)-\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)\right\}\right]^{i+1}}\right] \quad \text{i\_esima}$$

derivada , en consiguiente para $a=\theta=0$ y $b=1$ una aproximación de $f(1)$ es:

$$f(1)=f(0)-E\left[\frac{-\left(\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right)}{\left[\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)\right]^{2}}\right]-E\left[\frac{2\left(\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right)^{2}}{\left[\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)\right]^{3}}\right]-\ldots-E\left[\frac{(-1)^{i}i!\left(\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right)^{i}}{\left[\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)\right]^{i+1}}\right]$$

lo cual es equivalente a la siguiente expresión:

$$f(1)\cong E\left[\frac{\left(\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right)}{\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)^{2}}\right]-\sum_{i=1}^{\infty}E\left[\frac{(-1)^{i}\left(\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right)^{i}}{\left[\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)\right]^{i+1}}\right] \qquad (A.25)$$

en la cual suele considerarse como un valor insignificante el segundo término de la derecha de la expresión (A.25) con esto se obtiene la siguiente expresión:

$$E[b]\cong\frac{1}{\left(\sum_{i=1}^{N}X_{i}^{2}-N\bar{X}^{2}\right)^{2}}E\left[\sum_{i=1}^{n}x_{i}y_{i}-n\bar{x}\bar{y}\right]$$

aplicando el resultado (A.3) obtenemos lo siguiente:

$$E[b] \cong \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{\left(\sum_{i=1}^{N} X_i^2 - N\bar{X}^2\right)^2} = B \tag{A.26}$$

APÉNDICE B.1

EXPLICACIÓN A LOS CUADROS DE CONCENTRADOS DE RESULTADOS DE MEDIAS Y VARIANZAS MUESTRALES.

Cov: señala el valor de la correlación que existe entre la variable independiente y la variable dependiente.

N : indica el número de elementos de la población.

Media Poblacional : indica el valor verdadero de la media poblacional ($\bar{Y}$).

Varianza Poblacional: indica el valor verdadero de la variabilidad poblacional (S_y^2).

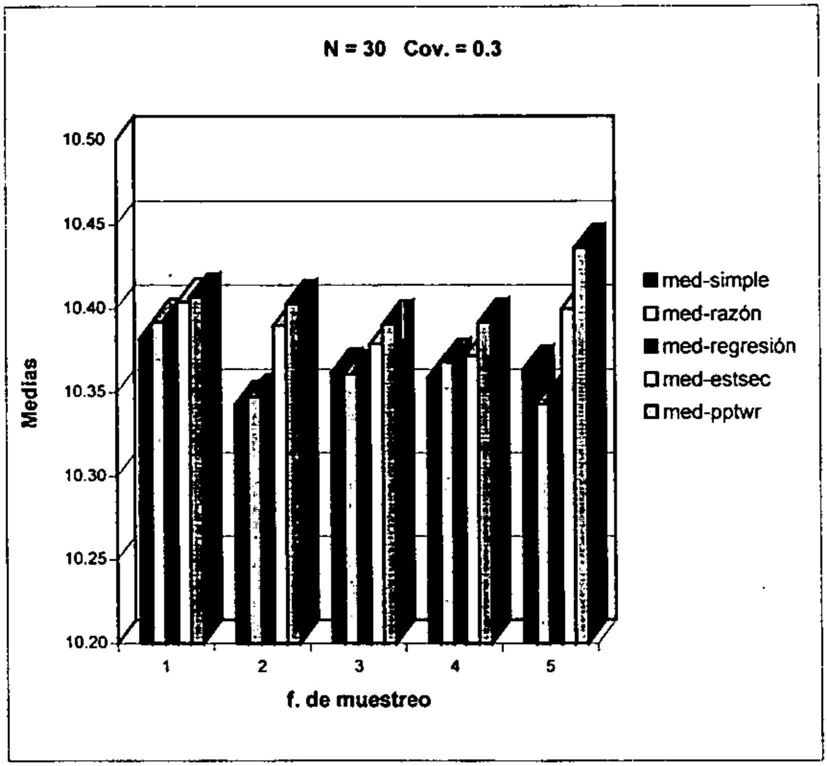
Media de Medias: cada valor, representa el promedio aritmético de los 200 valores evaluados, para un número semejante de muestras, de acuerdo a un estimador.

Varianza de Medias : cada valor, representa la variabilidad, que corresponde a los 200 valores evaluados de acuerdo con un estimador.

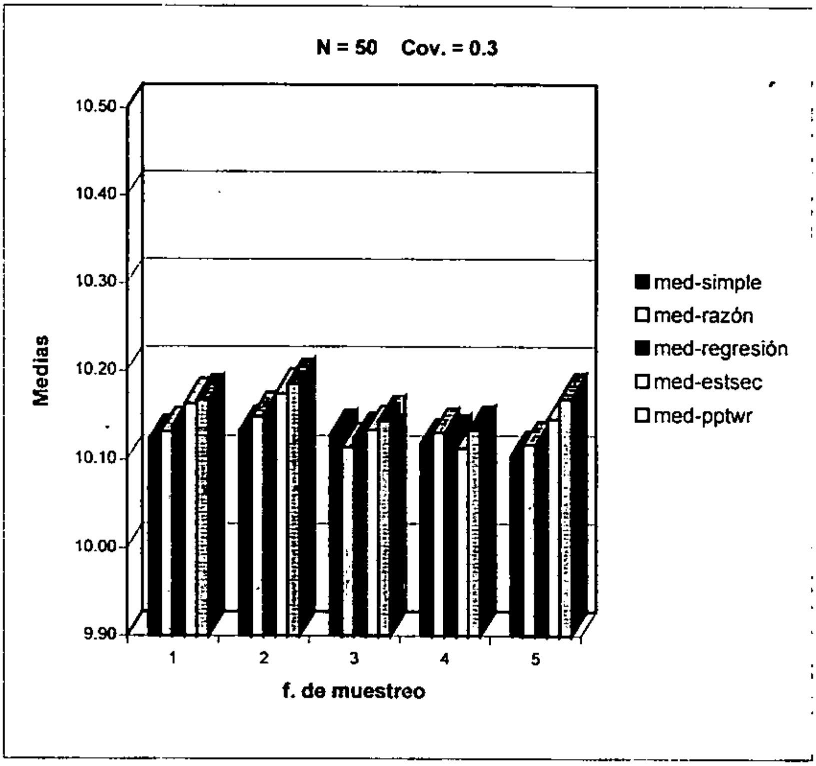
Media de Varianzas: cada valor indica el promedio aritmético de los 200 valores calculados, correspondientes a la estimación de la variabilidad de un estimador en cada una de las muestras realizadas.

F. de muestreo: es la fracción de muestreo que se utilizó.

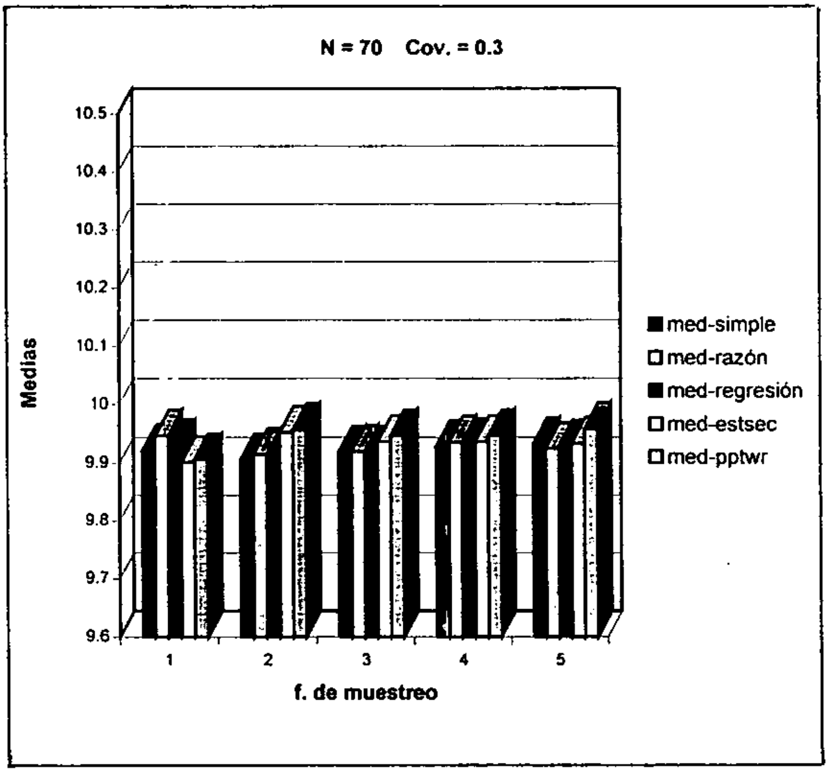
Media Poblacional = 10.37112					
N = 30			Cov. = 0.3		
f. de muestreo	1	2	3	4	5
med-simple	10.382	10.343	10.362	10.359	10.364
med-razón	10.392	10.347	10.361	10.368	10.343
med-regresión	10.394	10.333	10.357	10.358	10.351
med-estsec	10.404	10.390	10.379	10.372	10.400
med-pptwr	10.407	10.403	10.391	10.392	10.436



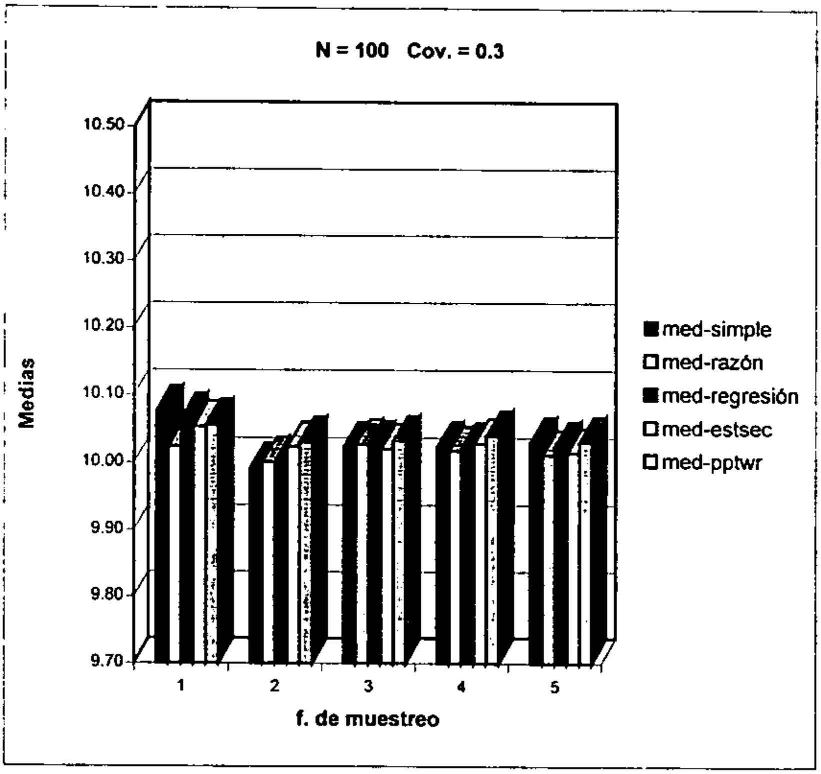
Media Poblacional = 10.11283					
N = 50			Cov. = 0.3		
f. de muestreo	1	2	3	4	5
med-simple	10.123	10.134	10.129	10.119	10.105
med-razón	10.132	10.149	10.114	10.131	10.117
med-regresión	10.139	10.151	10.126	10.124	10.108
med-estsec	10.164	10.175	10.134	10.113	10.147
med-pptwr	10.168	10.186	10.145	10.134	10.170



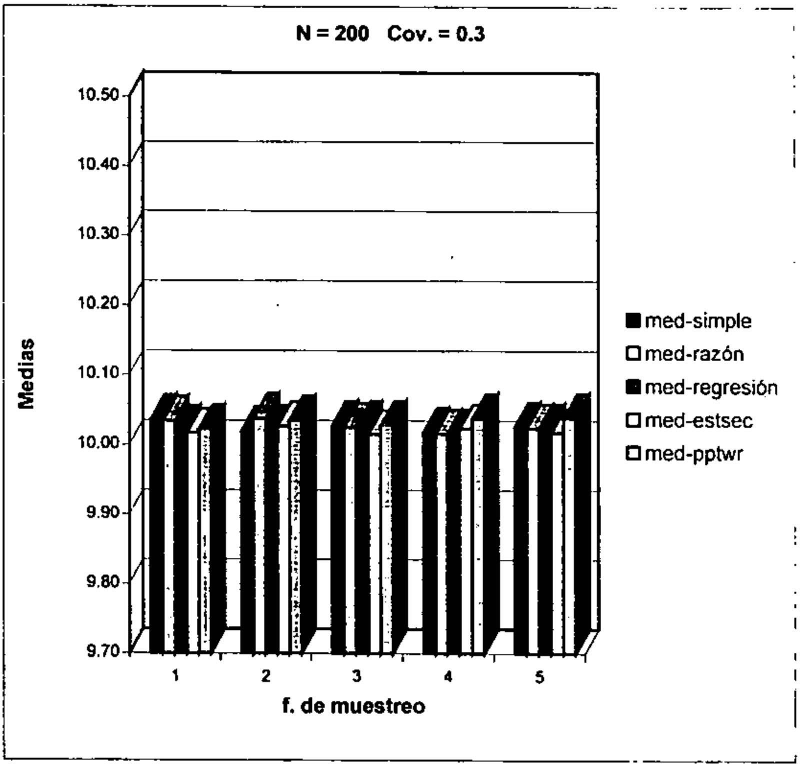
Media Poblacional = 9.93357					
N =70			Cov. = 0.3		
f. de muestreo	1	2	3	4	5
med-simple	9.922	9.906	9.919	9.929	9.933
med-razón	9.947	9.915	9.920	9.936	9.925
med-regresión	9.930	9.903	9.918	9.931	9.929
med-estsec	9.901	9.953	9.937	9.937	9.934
med-pptwr	9.906	9.957	9.947	9.947	9.959



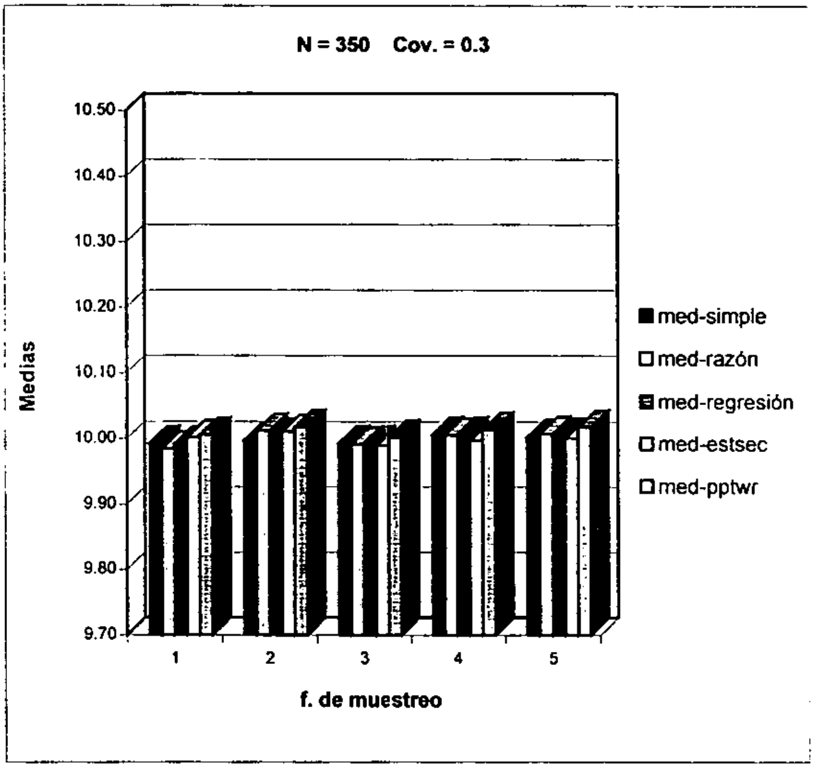
Media Poblacional = 10.02621					
	N =100		Cov. = 0.3		
f. de muestreo	1	2	3	4	5
med-simple	10.076	9.992	10.024	10.024	10.031
med-razón	10.024	10.001	10.028	10.018	10.012
med-regresión	10.065	9.993	10.025	10.023	10.025
med-estsec	10.054	10.024	10.021	10.029	10.015
med-pptwr	10.056	10.031	10.033	10.041	10.032



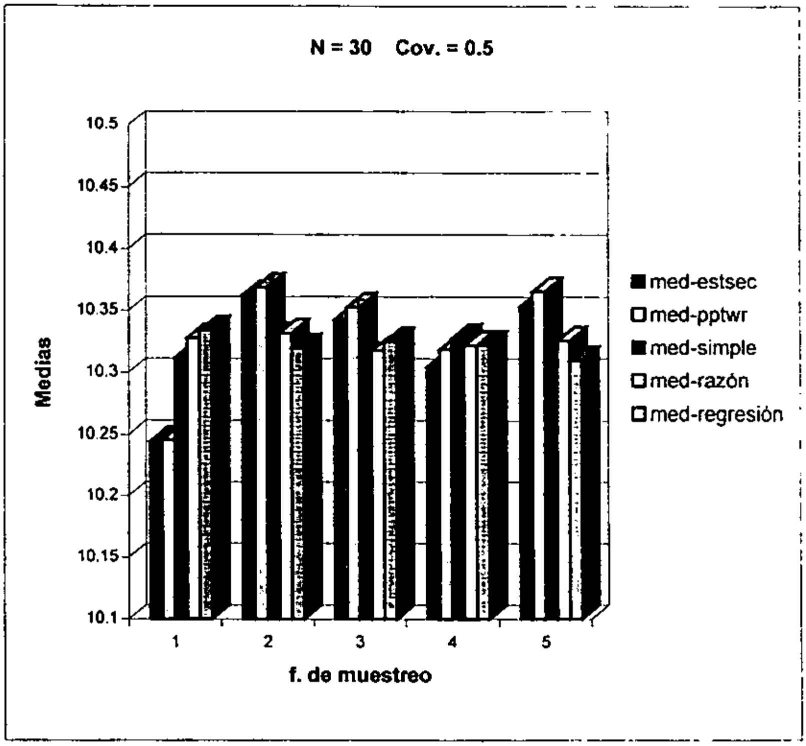
Media Poblacional = 10.02464					
N =200			Cov. = 0.3		
f. de muestreo	1	2	3	4	5
med-simple	10.036	10.020	10.028	10.019	10.027
med-razón	10.034	10.038	10.025	10.016	10.024
med-regresión	10.021	10.022	10.027	10.017	10.025
med-estsec	10.017	10.027	10.015	10.024	10.018
med-pptwr	10.022	10.035	10.028	10.038	10.039



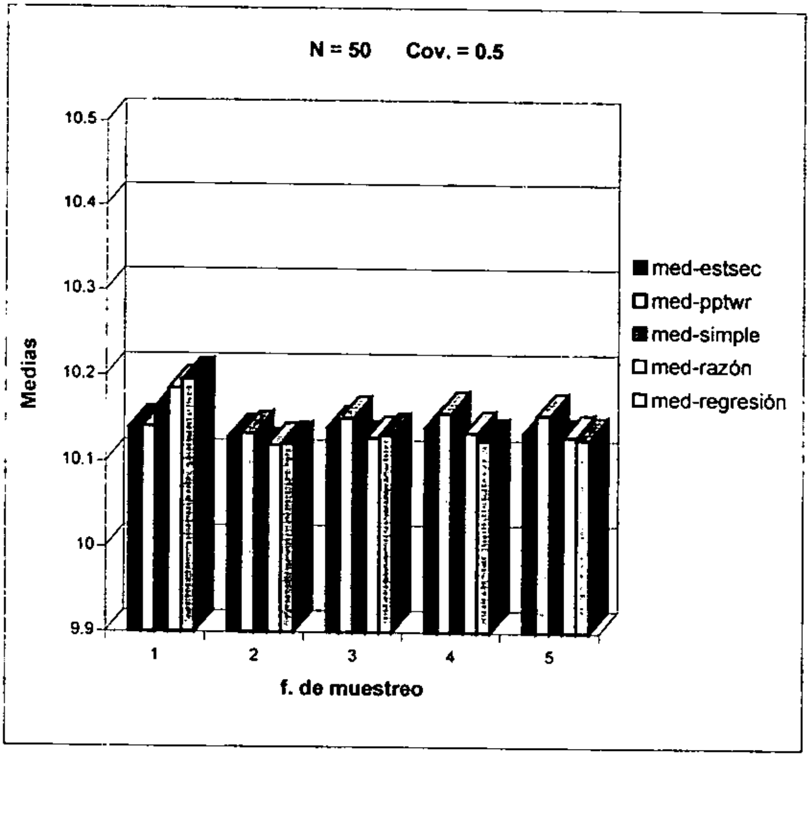
Media Poblacional = 9.99701					
N =350			Cov. = 0.3		
f. de muestreo	1	2	3	4	5
med-simple	9.993	9.998	9.993	10.007	10.003
med-razón	9.984	10.012	9.991	10.005	10.008
med-regresión	9.992	10.005	9.993	10.007	10.005
med-estsec	10.002	10.011	9.989	9.997	10.001
med-pptwr	10.006	10.018	10.002	10.014	10.018



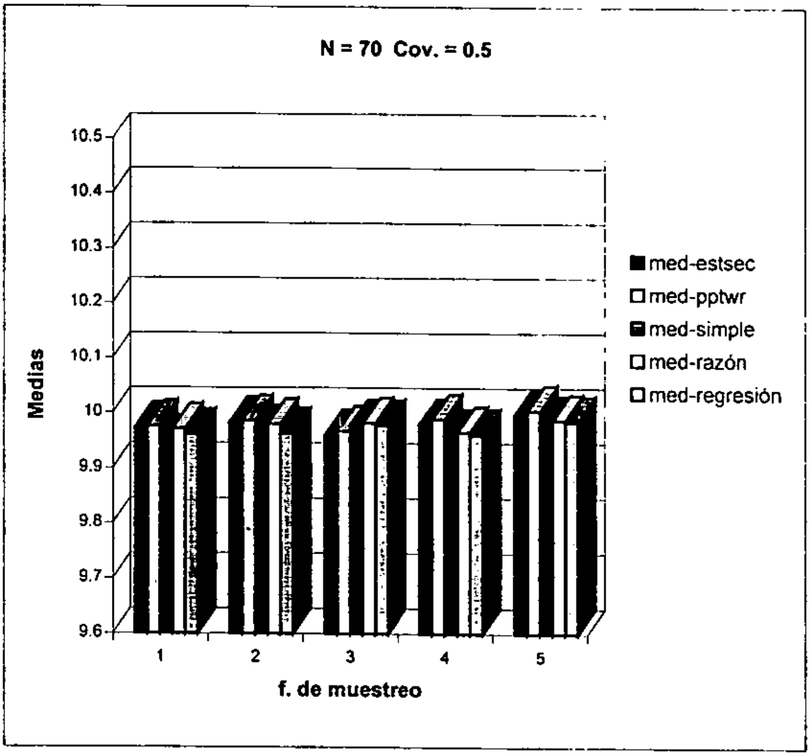
Media Poblacional = 10.33094					
N = 30			Cov. = 0.5		
f. de muestreo	1	2	3	4	5
med-simple	10.312	10.329	10.313	10.328	10.309
med-razón	10.328	10.332	10.318	10.322	10.326
med-regresión	10.334	10.320	10.325	10.322	10.310
med-estsec	10.245	10.363	10.343	10.305	10.354
med-pptwr	10.246	10.369	10.353	10.319	10.366



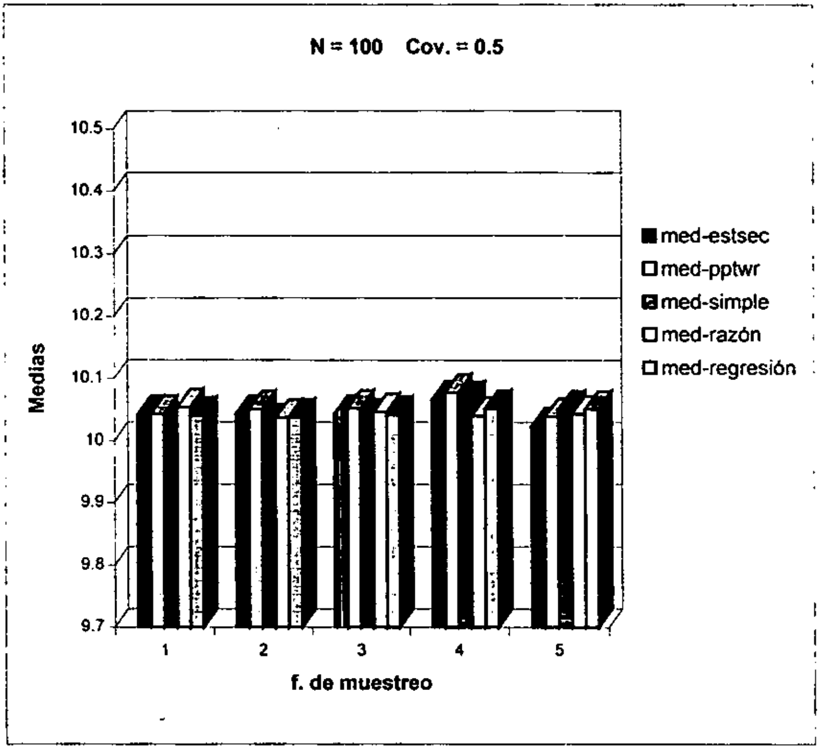
Media Poblacional = 10.12840					
N = 50			Cov. = 0.5		
f. de muestreo	1	2	3	4	5
med-simple	10.164	10.113	10.123	10.116	10.125
med-razón	10.185	10.120	10.128	10.134	10.130
med-regresión	10.195	10.121	10.131	10.125	10.127
med-estsec	10.140	10.129	10.140	10.141	10.135
med-pptwr	10.141	10.133	10.151	10.157	10.155



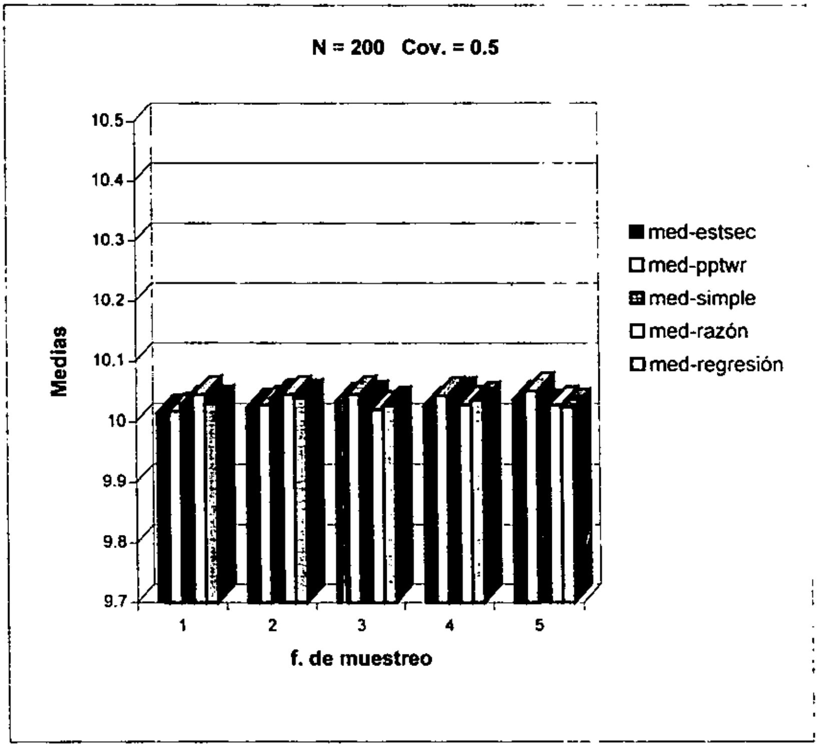
Media Poblacional = 9.97667					
N = 70			Cov. = 0.5		
f. de muestreo	1	2	3	4	5
med-simple	9.949	9.968	9.978	9.962	9.986
med-razón	9.971	9.980	9.982	9.966	9.988
med-regresión	9.961	9.963	9.978	9.961	9.985
med-estsec	9.973	9.983	9.960	9.981	9.999
med-pptwr	9.976	9.986	9.967	9.989	10.005



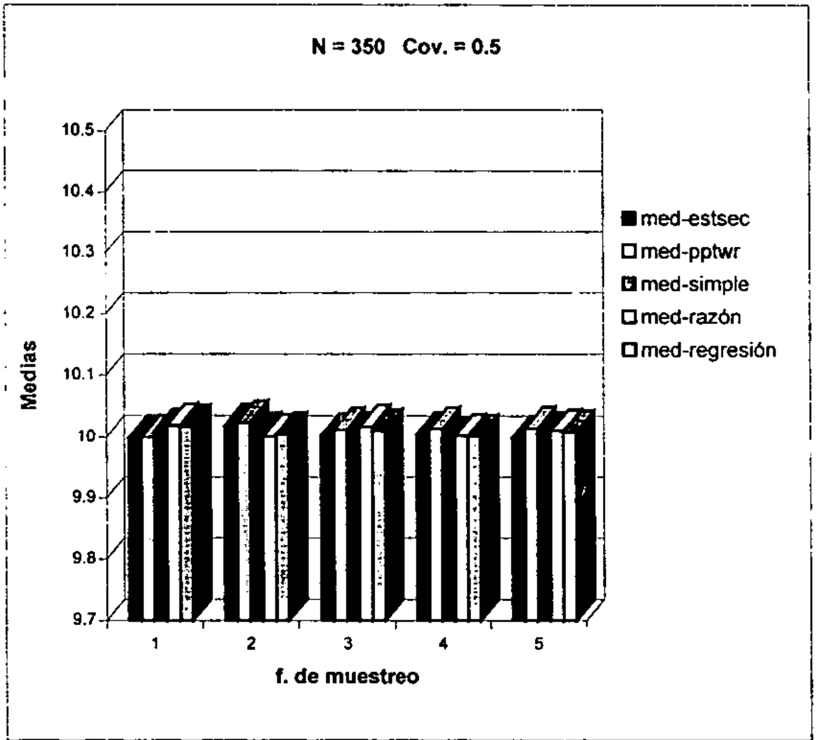
Media Poblacional = 10.04616					
N = 100			Cov. = 0.5		
f. de muestreo	1	2	3	4	5
med-simple	10.015	10.028	10.037	10.065	10.050
med-razón	10.053	10.036	10.046	10.039	10.042
med-regresión	10.040	10.037	10.041	10.051	10.050
med-estsec	10.042	10.043	10.043	10.066	10.023
med-pptwr	10.042	10.050	10.051	10.077	10.038



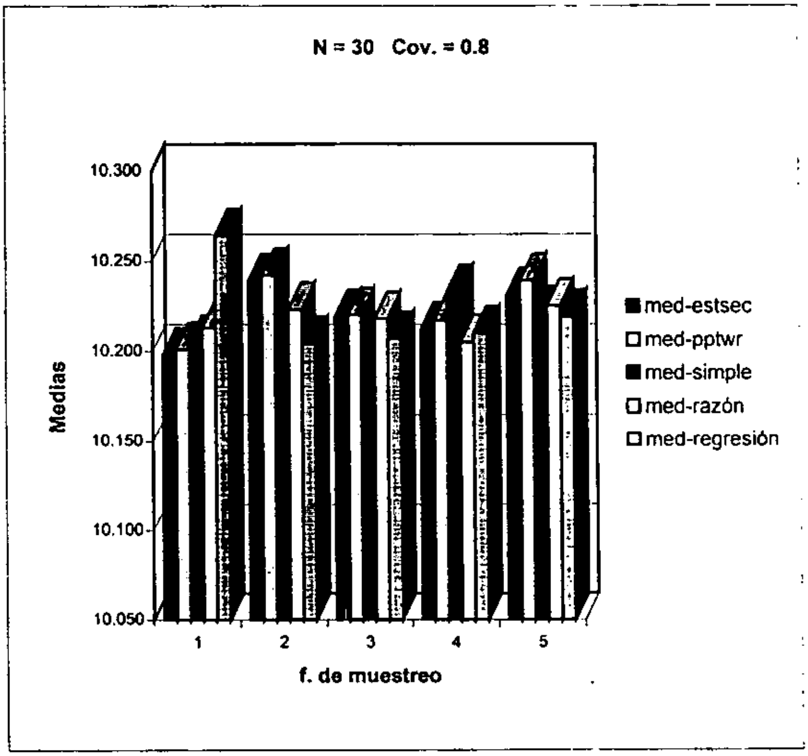
Media Poblacional = 10.03317					
N = 200			Cov. = 0.5		
f. de muestreo	1	2	3	4	5
med-simple	10.029	10.039	10.036	10.042	10.022
med-razón	10.045	10.045	10.020	10.028	10.028
med-regresión	10.029	10.039	10.027	10.035	10.024
med-estsec	10.015	10.024	10.034	10.027	10.037
med-pptwr	10.017	10.028	10.045	10.043	10.051



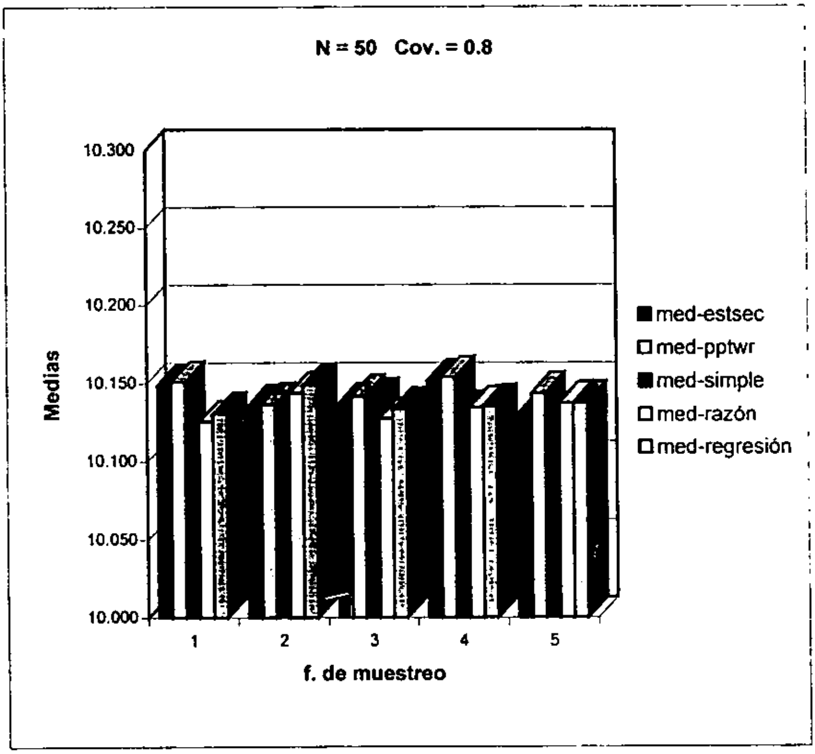
Media Poblacional = 9.99701					
N = 350			Cov. = 0.5		
f. de muestreo	1	2	3	4	5
med-simple	10.010	10.004	10.005	9.998	10.006
med-razón	10.019	10.001	10.017	10.002	10.010
med-regresión	10.017	10.004	10.010	10.001	10.008
med-estsec	9.999	10.019	10.003	10.005	10.000
med-pptwr	10.000	10.024	10.012	10.013	10.014



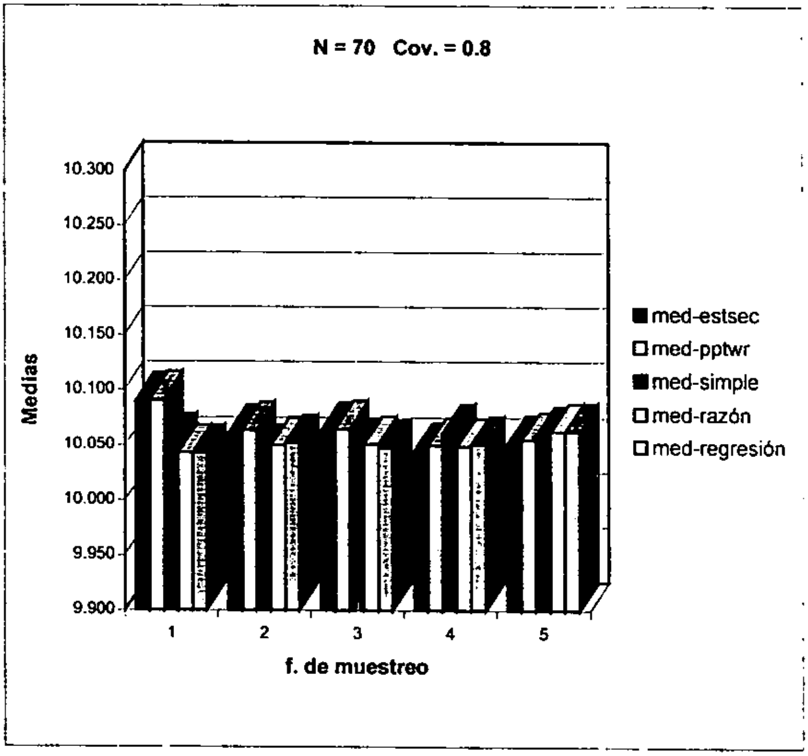
Media Poblacional = 10.21736					
N = 30			Cov. = 0.8		
f. de muestreo	1	2	3	4	5
med-simple	10.205	10.129	10.206	10.233	10.221
med-razón	10.213	10.223	10.218	10.205	10.225
med-regresión	10.264	10.204	10.207	10.210	10.219
med-estsec	10.198	10.239	10.219	10.212	10.231
med-pptwr	10.201	10.242	10.220	10.217	10.239



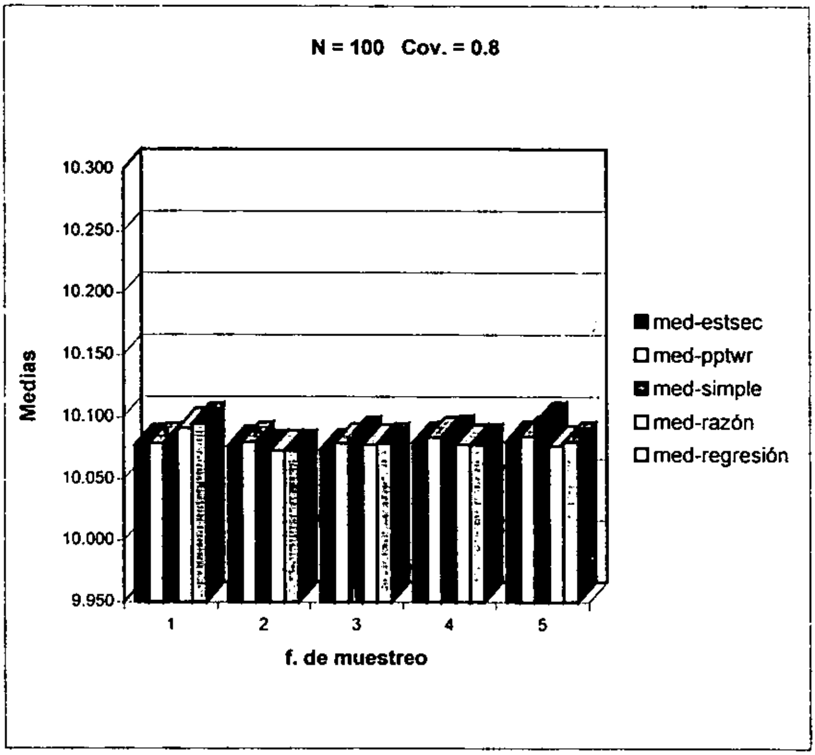
Media Poblacional = 10.14071					
N = 50			Cov. = 0.8		
f. de muestreo	1	2	3	4	5
med-simple	10.114	10.134	10.140	10.132	10.130
med-razón	10.126	10.144	10.128	10.135	10.138
med-regresión	10.131	10.149	10.134	10.136	10.138
med-estsec	10.149	10.136	10.137	10.152	10.131
med-pptwr	10.151	10.137	10.142	10.154	10.144



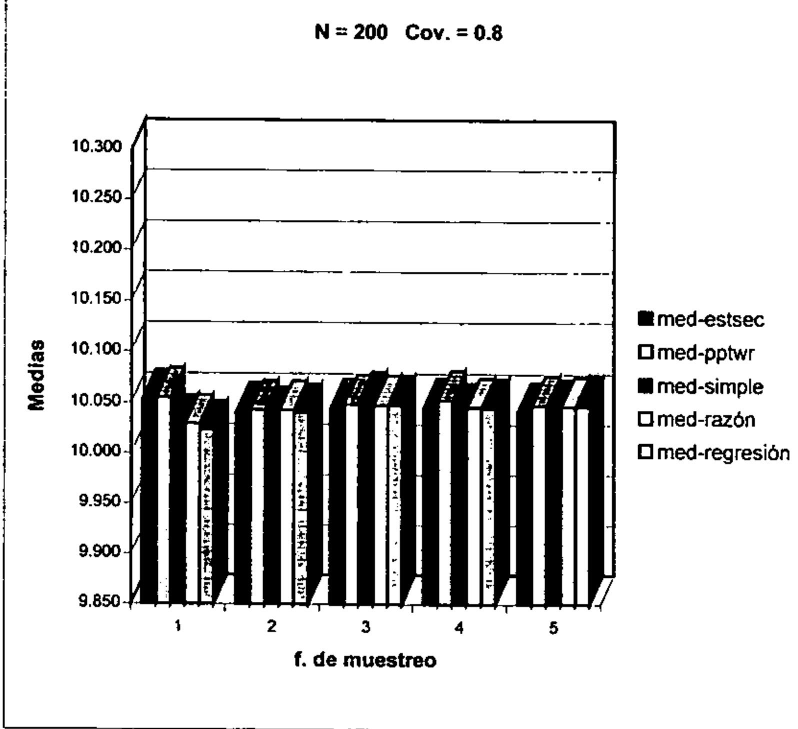
Media Poblacional = 10.05752					
	N = 70		Cov. = 0.8		
f. de muestreo	1	2	3	4	5
med-simple	10.053	10.043	10.049	10.062	10.060
med-razón	10.043	10.050	10.051	10.049	10.063
med-regresión	10.043	10.052	10.048	10.051	10.063
med-estsec	10.090	10.061	10.064	10.045	10.052
med-pptwr	10.091	10.064	10.065	10.050	10.055



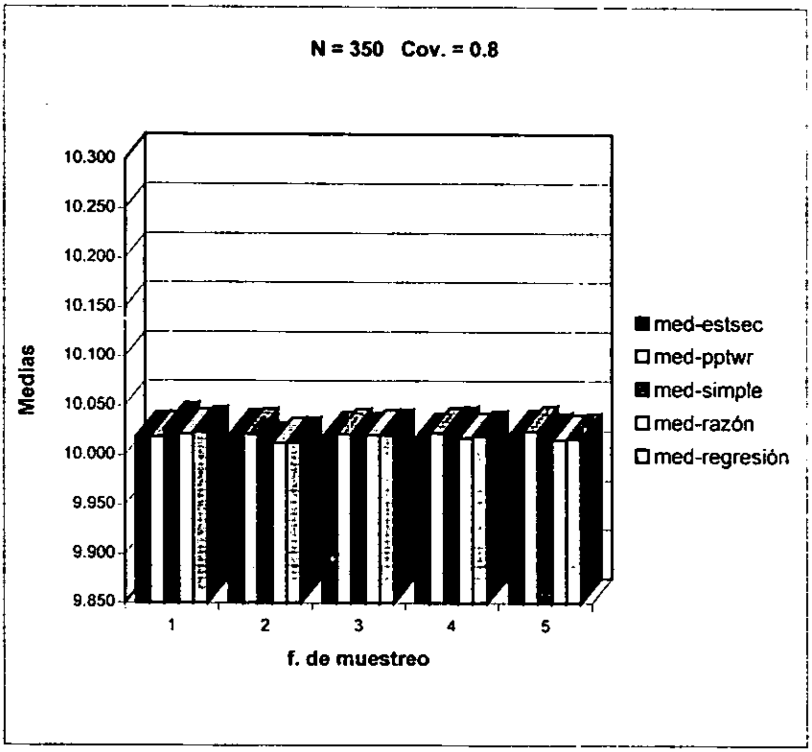
Media Poblacional = 10.07654					
N = 100			Cov. = 0.8		
f. de muestreo	1	2	3	4	5
med-simple	10.075	10.073	10.084	10.085	10.096
med-razón	10.091	10.073	10.078	10.078	10.077
med-regresión	10.095	10.073	10.079	10.078	10.080
med-estsec	10.078	10.077	10.074	10.080	10.081
med-pptwr	10.079	10.080	10.079	10.084	10.085



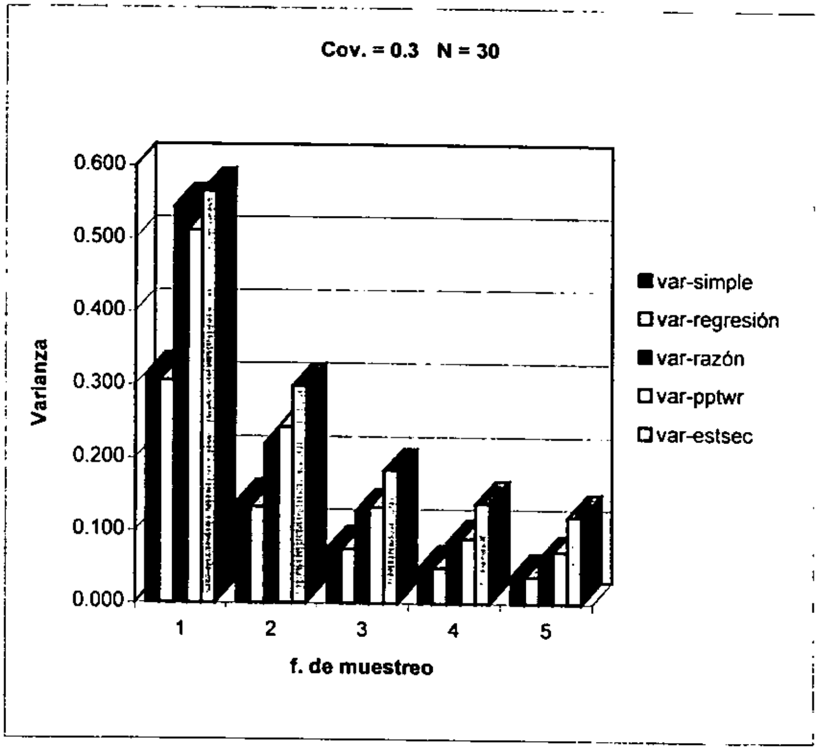
Media Poblacional = 10.04448					
N = 200			Cov. = 0.8		
f. de muestreo	1	2	3	4	5
med-simple	10.029	10.037	10.052	10.041	10.045
med-razón	10.029	10.043	10.048	10.046	10.047
med-regresión	10.023	10.040	10.048	10.045	10.047
med-estsec	10.053	10.041	10.045	10.046	10.042
med-pptwr	10.055	10.043	10.049	10.053	10.048



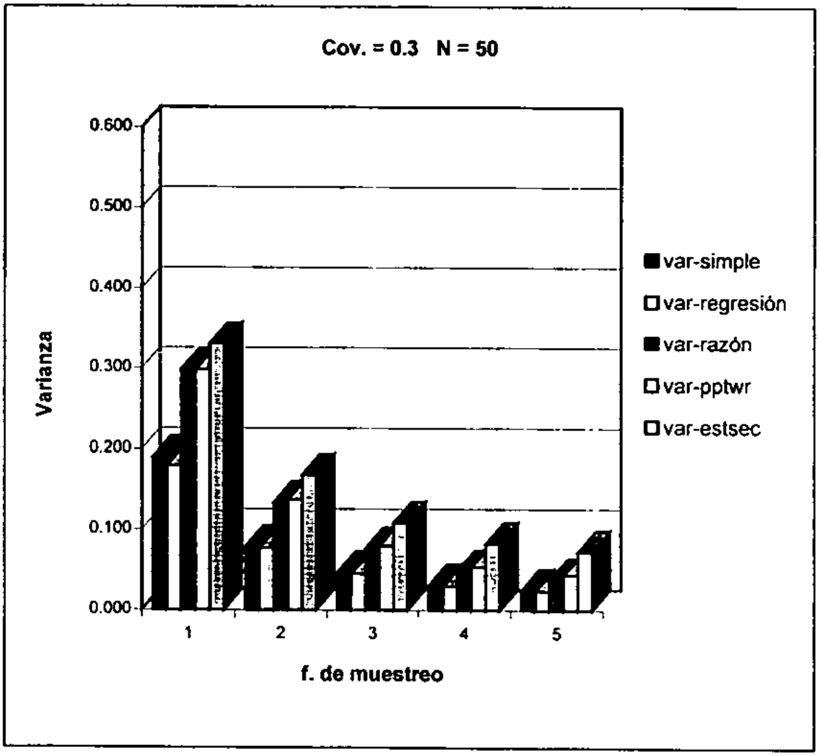
Media Poblacional = 10.01905					
	N = 350		Cov. = 0.8		
f. de muestreo	1	2	3	4	5
med-simple	10.028	10.008	10.018	10.024	10.018
med-razón	10.022	10.013	10.021	10.018	10.016
med-regresión	10.024	10.013	10.021	10.020	10.017
med-estsec	10.019	10.021	10.019	10.018	10.021
med-pptwr	10.019	10.022	10.022	10.023	10.025



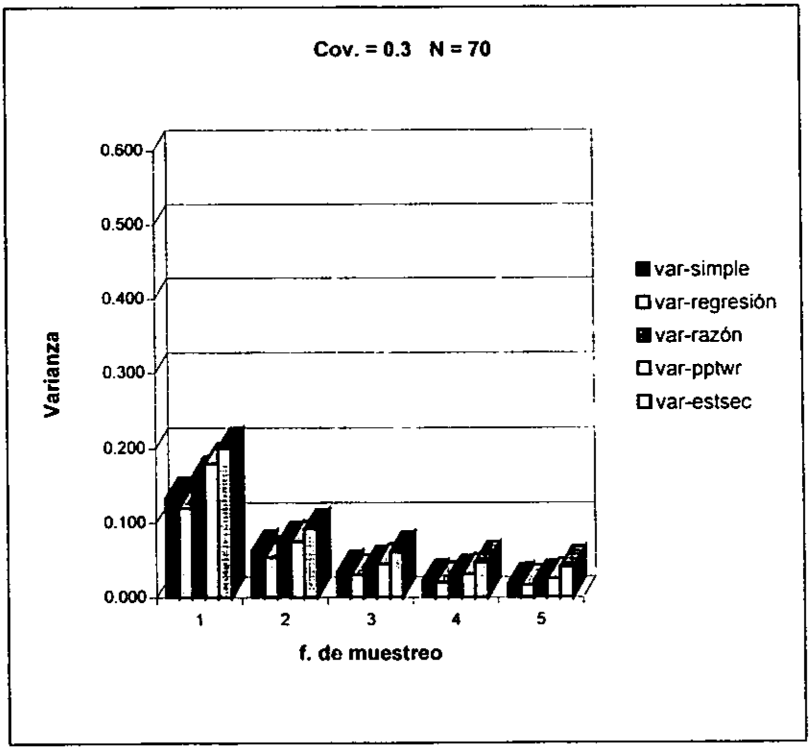
Varianza Poblacional = 1.05109					
Cov.= 0.3			N = 30		
f. de muestreo	1	2	3	4	5
var-simple	0.314	0.141	0.077	0.050	0.039
var-regresión	0.306	0.133	0.075	0.049	0.037
var-razón	0.545	0.221	0.128	0.086	0.064
var-pptwr	0.512	0.243	0.133	0.090	0.072
var-estsec	0.566	0.299	0.184	0.139	0.121



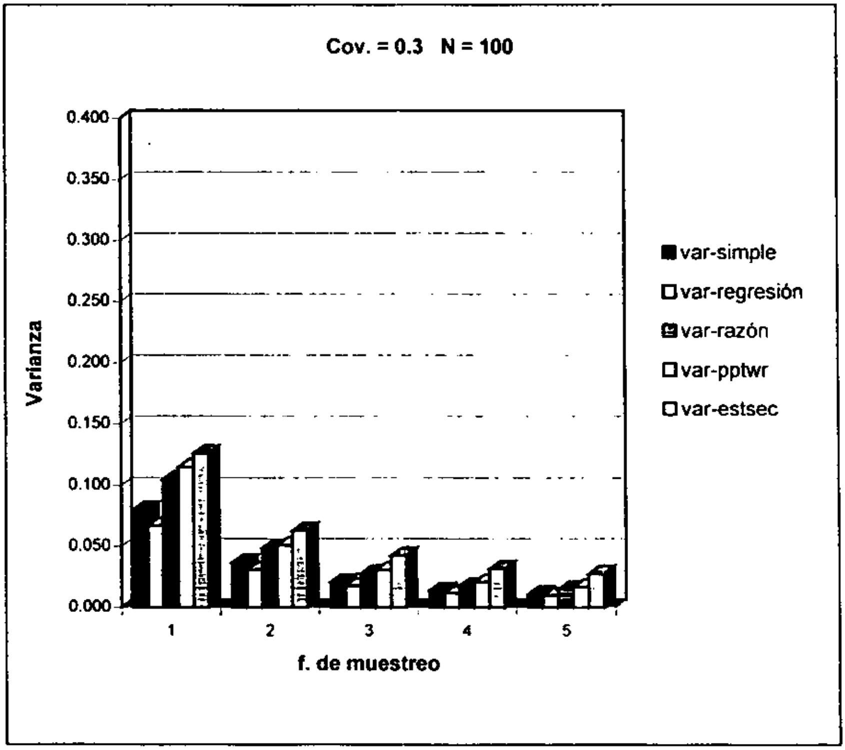
Varianza Poblacional = 1.04682					
Cov. = 0.3			N = 50		
f. de muestreo	1	2	3	4	5
var-simple	0.190	0.079	0.047	0.031	0.025
var-regresión	0.179	0.077	0.045	0.030	0.024
var-razón	0.299	0.134	0.077	0.049	0.038
var-pptwr	0.298	0.137	0.080	0.054	0.044
var-estsec	0.331	0.168	0.109	0.083	0.073



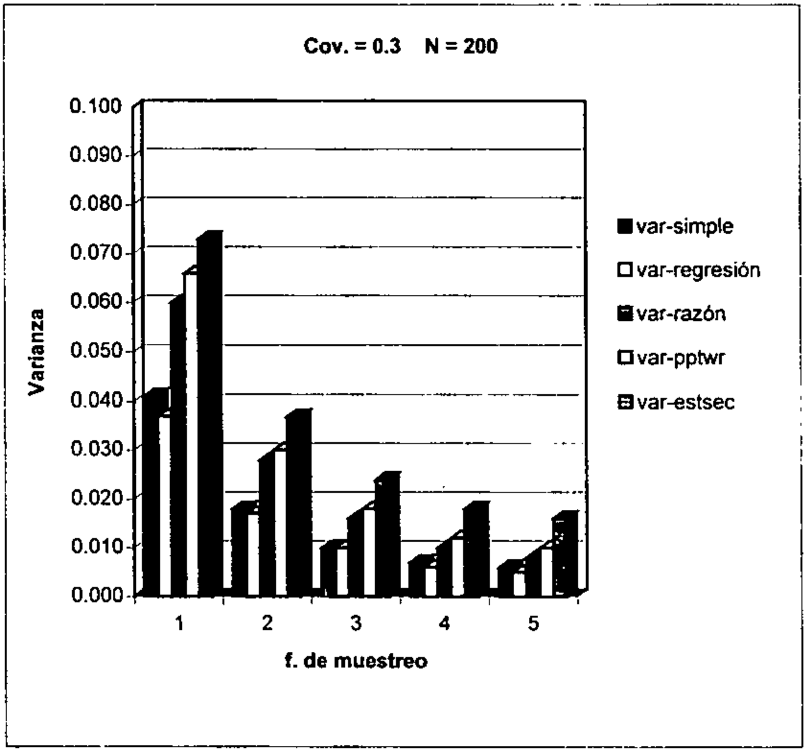
Varianza Poblacional = 1.07218					
Cov. = 03			N = 70		
f. de muestreo	1	2	3	4	5
var-simple	0.134	0.062	0.035	0.023	0.018
var-regresión	0.120	0.054	0.030	0.020	0.016
var-razón	0.163	0.074	0.042	0.029	0.022
var-pptwr	0.180	0.075	0.045	0.031	0.025
var-estsec	0.200	0.092	0.061	0.047	0.041



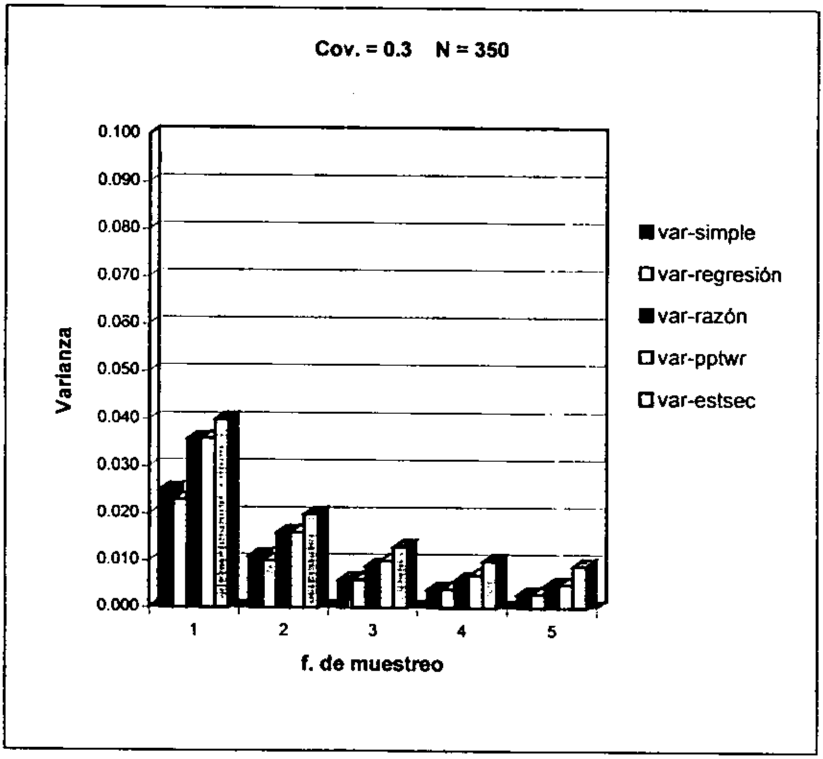
Varianza Poblacional = 0.92289					
Cov. = 0.3			N = 100		
f. de muestreo	1	2	3	4	5
var-simple	0.080	0.037	0.021	0.014	0.011
var-regresión	0.067	0.031	0.018	0.012	0.010
var-razón	0.105	0.049	0.028	0.018	0.015
var-pptwr	0.115	0.051	0.031	0.021	0.017
var-estsec	0.126	0.063	0.043	0.032	0.028



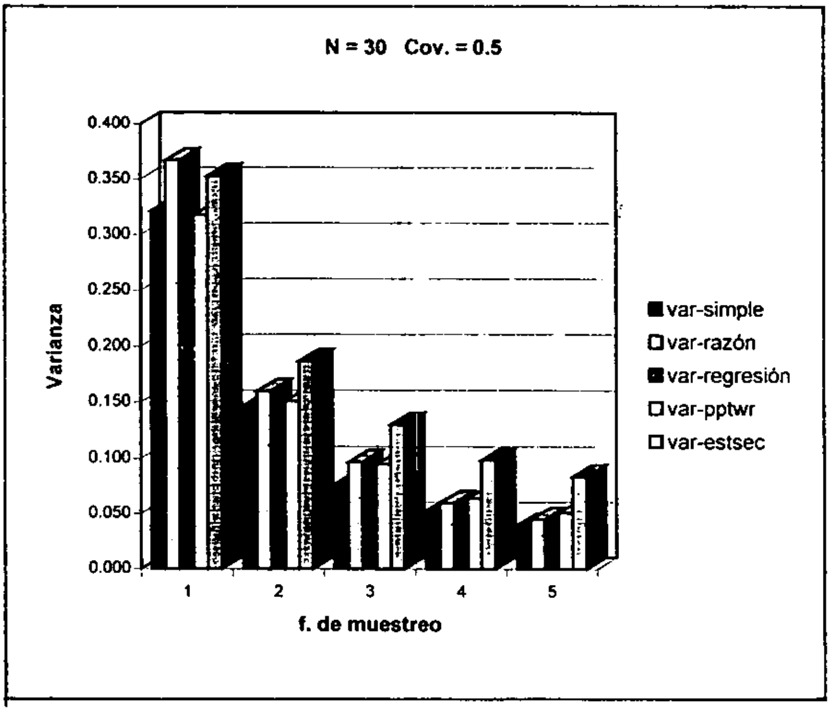
Varianza Poblacional = 0.90752					
Cov. = 0.3			N = 200		
f. de muestreo	1	2	3	4	5
var-simple	0.041	0.018	0.010	0.007	0.006
var-regresión	0.037	0.017	0.010	0.006	0.005
var-razón	0.060	0.028	0.016	0.010	0.008
var-pptwr	0.066	0.030	0.018	0.012	0.010
var-estsec	0.073	0.037	0.024	0.018	0.016



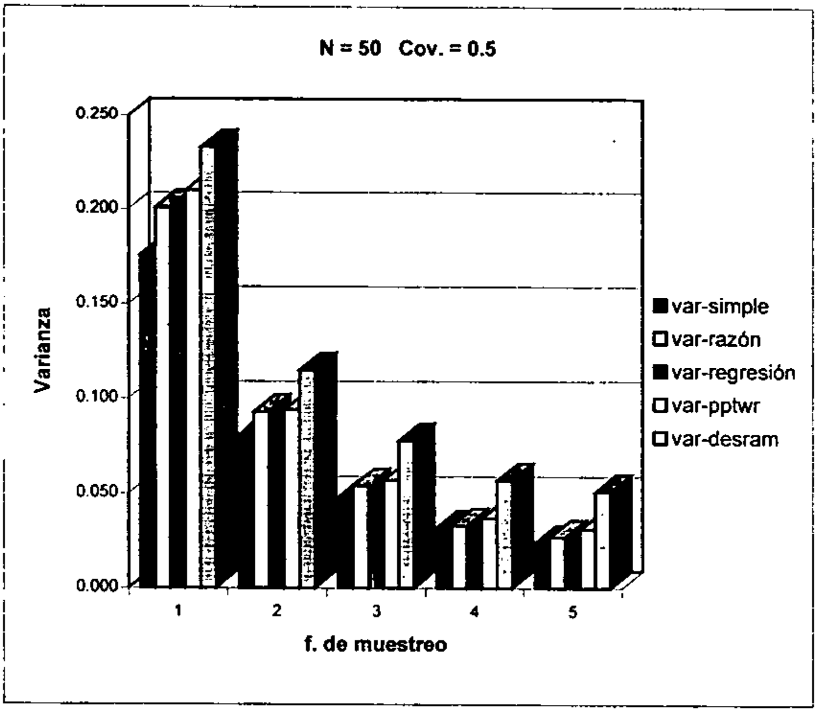
Varianza Poblacional = 0.96897					
Cov. = 0.3			N = 350		
f. de muestreo	1	2	3	4	5
var-simple	0.025	0.011	0.006	0.004	0.003
var-regresión	0.023	0.010	0.006	0.004	0.003
var-razón	0.036	0.016	0.009	0.006	0.005
var-pptwr	0.036	0.016	0.010	0.007	0.005
var-estsec	0.040	0.020	0.013	0.010	0.009



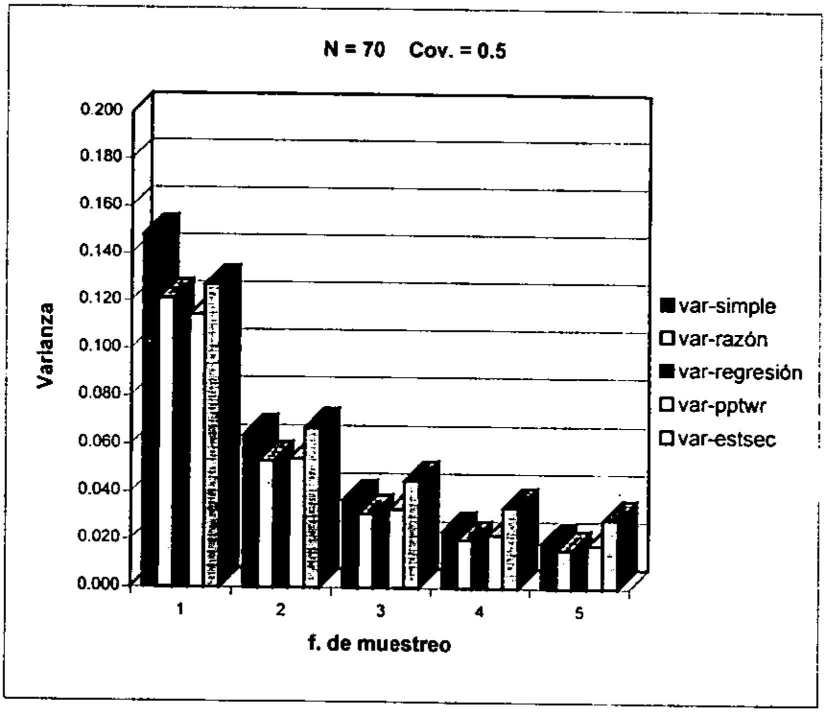
Varianza Poblacional = 1.06690					
Cov. = 0.5			N = 30		
f. de muestreo	1	2	3	4	5
var-simple	0.322	0.147	0.077	0.053	0.040
var-razón	0.367	0.160	0.097	0.060	0.045
var-regresión	0.257	0.112	0.065	0.041	0.031
var-pptwr	0.318	0.151	0.095	0.064	0.051
var-estsec	0.353	0.187	0.130	0.099	0.084



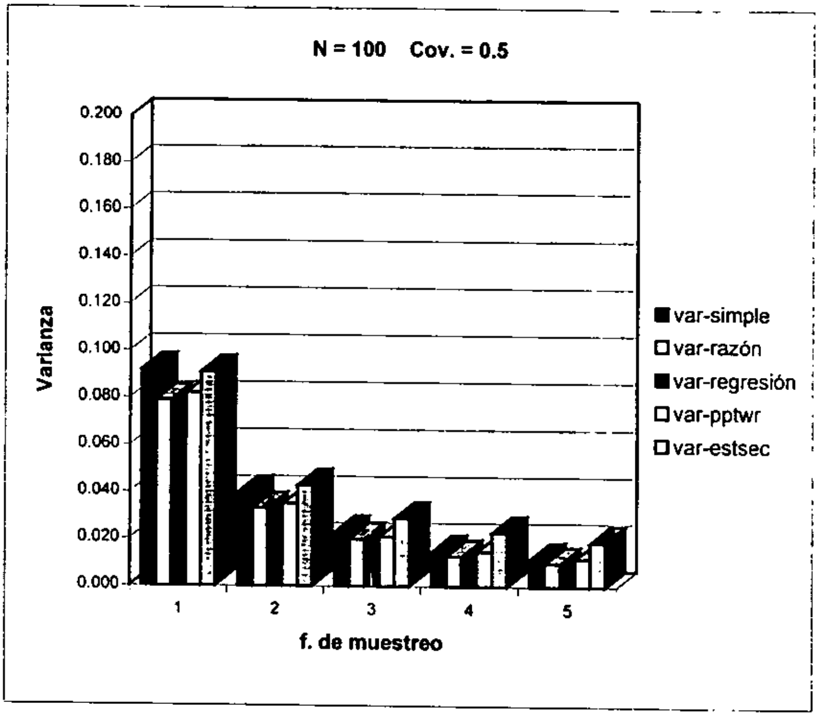
Varianza Poblacional = 1.04557					
Cov. = 0.5			N = 50		
f. de muestreo	1	2	3	4	5
var-simple	0.176	0.081	0.048	0.032	0.024
var-razón	0.201	0.093	0.054	0.033	0.027
var-regresión	0.139	0.064	0.039	0.024	0.019
var-pptwr	0.210	0.094	0.057	0.037	0.031
var-desram	0.233	0.115	0.078	0.057	0.051



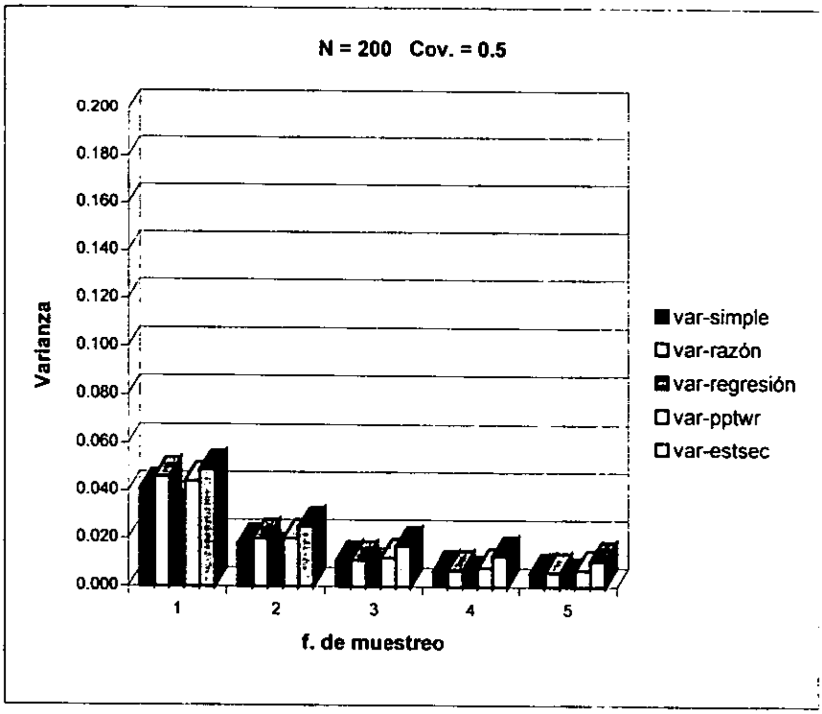
Varianza Poblacional = 1.10787					
Cov. = 0.5			N = 70		
f. de muestreo	1	2	3	4	5
var-simple	0.148	0.064	0.037	0.024	0.019
var-razón	0.121	0.053	0.031	0.020	0.016
var-regresión	0.104	0.043	0.025	0.016	0.013
var-pptwr	0.114	0.054	0.033	0.022	0.018
var-estsec	0.127	0.067	0.045	0.034	0.029



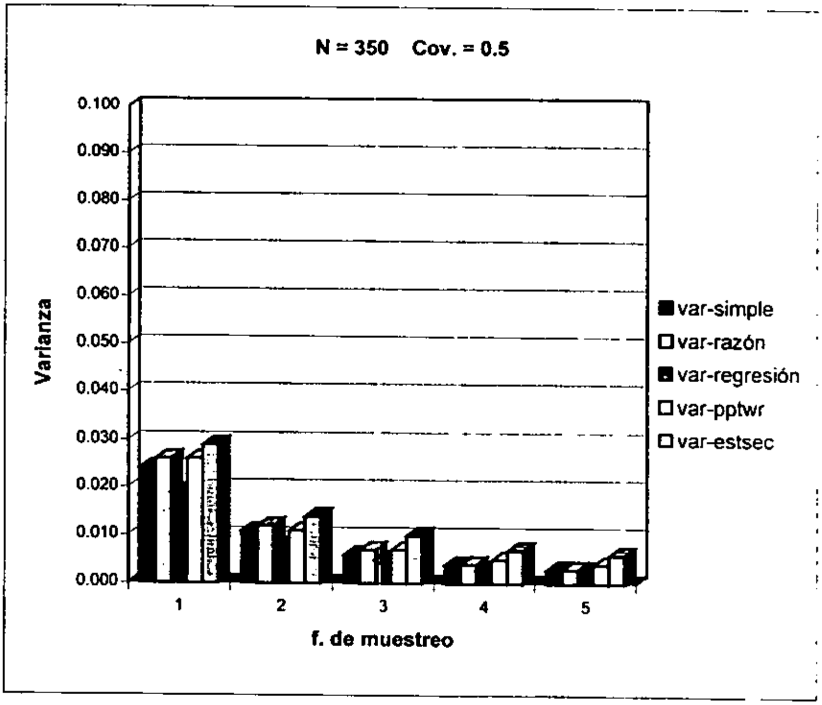
Varianza Poblacional = 0.98660					
Cov. = 0.5			N = 100		
f. de muestreo	1	2	3	4	5
var-simple	0.092	0.040	0.023	0.015	0.012
var-razón	0.079	0.033	0.020	0.013	0.010
var-regresión	0.062	0.026	0.015	0.010	0.008
var-pptwr	0.082	0.035	0.021	0.015	0.012
var-estsec	0.091	0.043	0.029	0.023	0.019



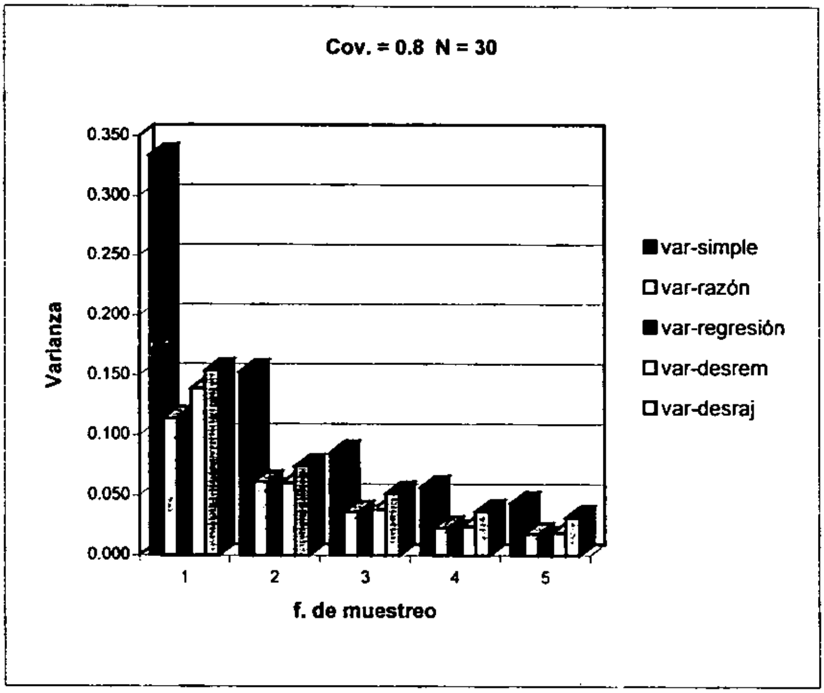
Varianza Poblacional = 0.90171					
Cov. = 0.5			N = 200		
f. de muestreo	1	2	3	4	5
var-simple	0.041	0.018	0.011	0.007	0.006
var-razón	0.046	0.020	0.011	0.007	0.006
var-regresión	0.032	0.014	0.008	0.005	0.004
var-pptwr	0.044	0.020	0.012	0.008	0.007
var-estsec	0.049	0.025	0.017	0.013	0.011



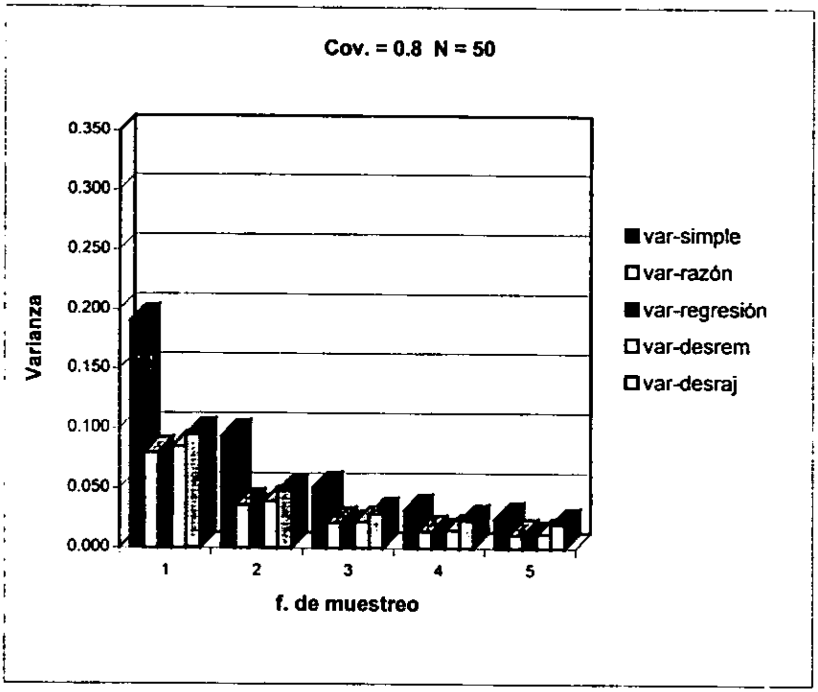
Varianza Poblacional = 0.95367					
Cov. = 0.5			N = 350		
f. de muestreo	1	2	3	4	5
var-simple	0.024	0.011	0.006	0.004	0.003
var-razón	0.026	0.012	0.007	0.004	0.003
var-regresión	0.019	0.008	0.005	0.003	0.003
var-pptwr	0.026	0.011	0.007	0.005	0.004
var-estsec	0.029	0.014	0.010	0.007	0.006



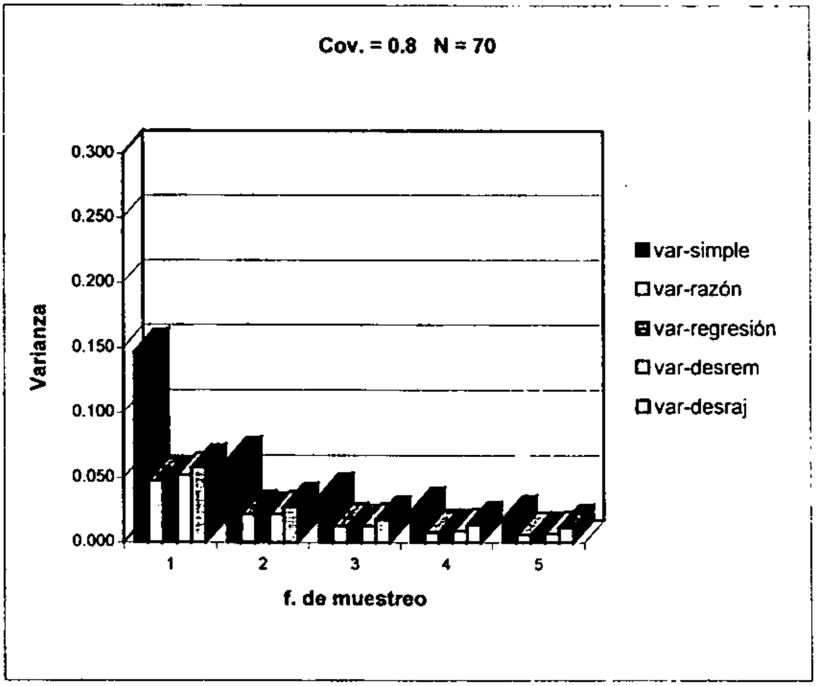
Varianza Poblacional = 1.12949					
Cov. = 0.8			N = 30		
f. de muestreo	1	2	3	4	5
var-simple	0.334	0.153	0.086	0.057	0.044
var-razón	0.114	0.061	0.036	0.023	0.018
var-regresión	0.098	0.054	0.032	0.020	0.015
var-desrem	0.139	0.060	0.038	0.024	0.019
var-desraj	0.154	0.075	0.052	0.037	0.032



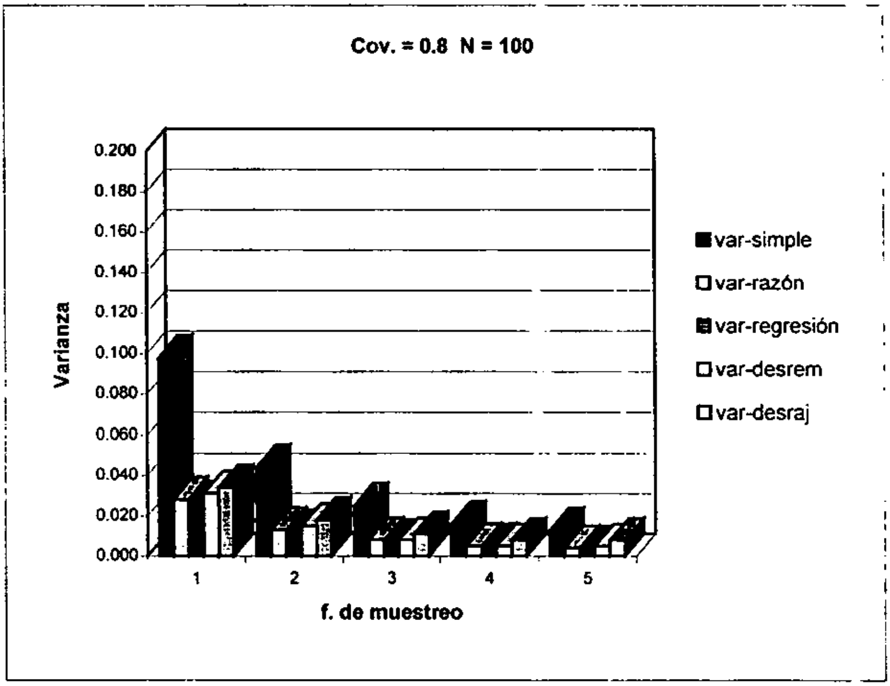
Varianza Poblacional = 1.08076					
Cov. = 0.8			N = 50		
f. de muestreo	1	2	3	4	5
var-simple	0.191	0.094	0.051	0.032	0.025
var-razón	0.080	0.036	0.021	0.014	0.011
var-regresión	0.070	0.032	0.018	0.012	0.009
var-desrem	0.085	0.039	0.022	0.015	0.012
var-desraj	0.095	0.048	0.029	0.023	0.020



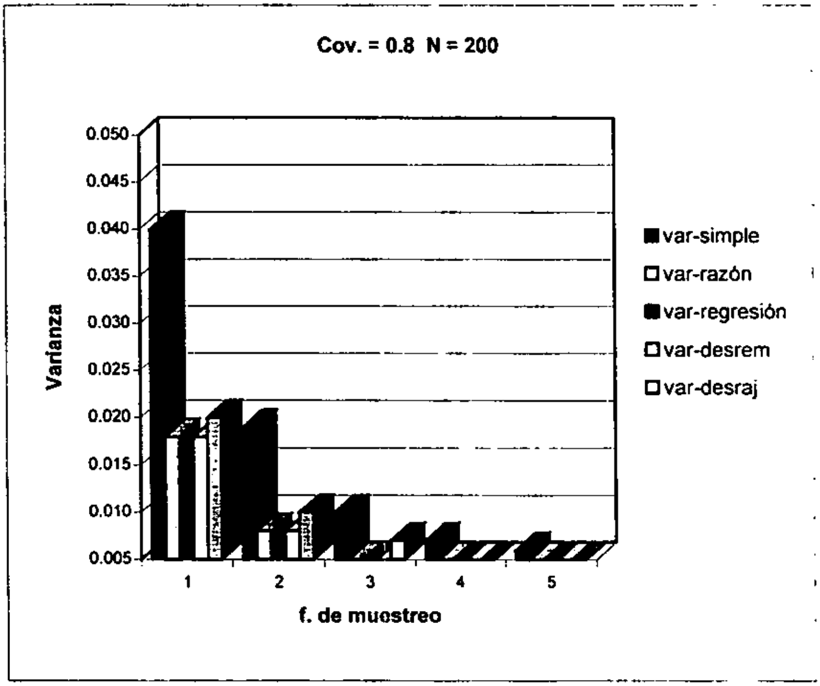
Varianza Poblacional = 1.13884					
Cov. = 0.8			N = 70		
f. de muestreo	1	2	3	4	5
var-simple	0.147	0.064	0.036	0.025	0.019
var-razón	0.048	0.022	0.013	0.008	0.006
var-regresión	0.048	0.021	0.012	0.008	0.006
var-desrem	0.052	0.022	0.013	0.009	0.007
var-desraj	0.058	0.028	0.018	0.014	0.012



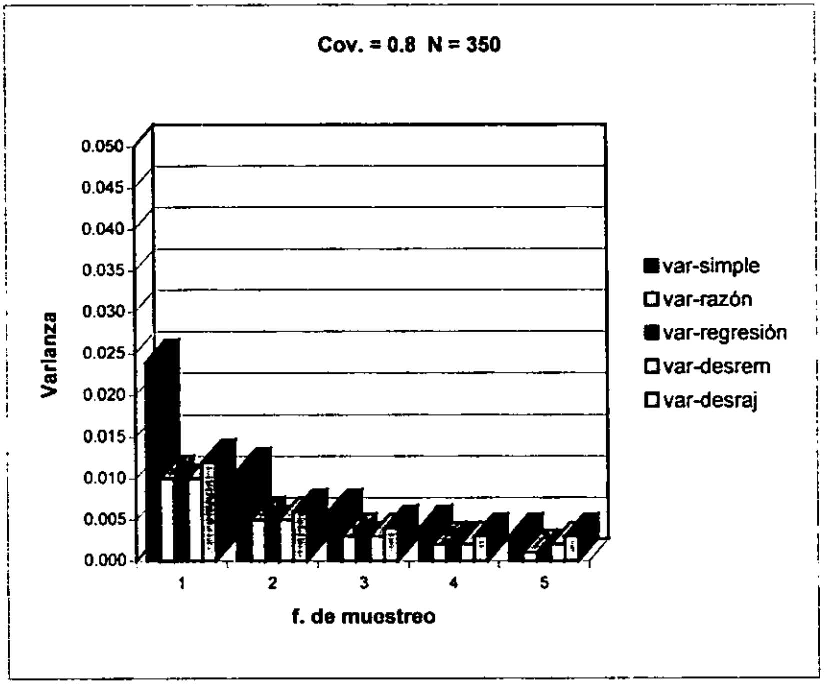
Varianza Poblacional = 1.09441					
Cov. = 0.8			N = 100		
f. de muestreo	1	2	3	4	5
var-simple	0.098	0.044	0.025	0.016	0.013
var-razón	0.028	0.013	0.008	0.005	0.004
var-regresión	0.026	0.012	0.007	0.005	0.004
var-desrem	0.031	0.015	0.008	0.005	0.005
var-desraj	0.034	0.018	0.011	0.008	0.008



Varianza Poblacional = 0.91755					
Cov. = 0.8			N = 200		
f. de muestreo	1	2	3	4	5
var-simple	0.040	0.019	0.010	0.007	0.006
var-razón	0.018	0.008	0.004	0.003	0.002
var-regresión	0.016	0.007	0.004	0.003	0.002
var-desrem	0.018	0.008	0.005	0.003	0.003
var-desraj	0.020	0.010	0.007	0.005	0.004



Varianza Poblacional = 0.93606					
Cov. = 0.8			N = 350		
f. de muestreo	1	2	3	4	5
var-simple	0.024	0.011	0.006	0.004	0.003
var-razón	0.010	0.005	0.003	0.002	0.001
var-regresión	0.009	0.004	0.002	0.002	0.001
var-desrem	0.010	0.005	0.003	0.002	0.002
var-desraj	0.012	0.006	0.004	0.003	0.003



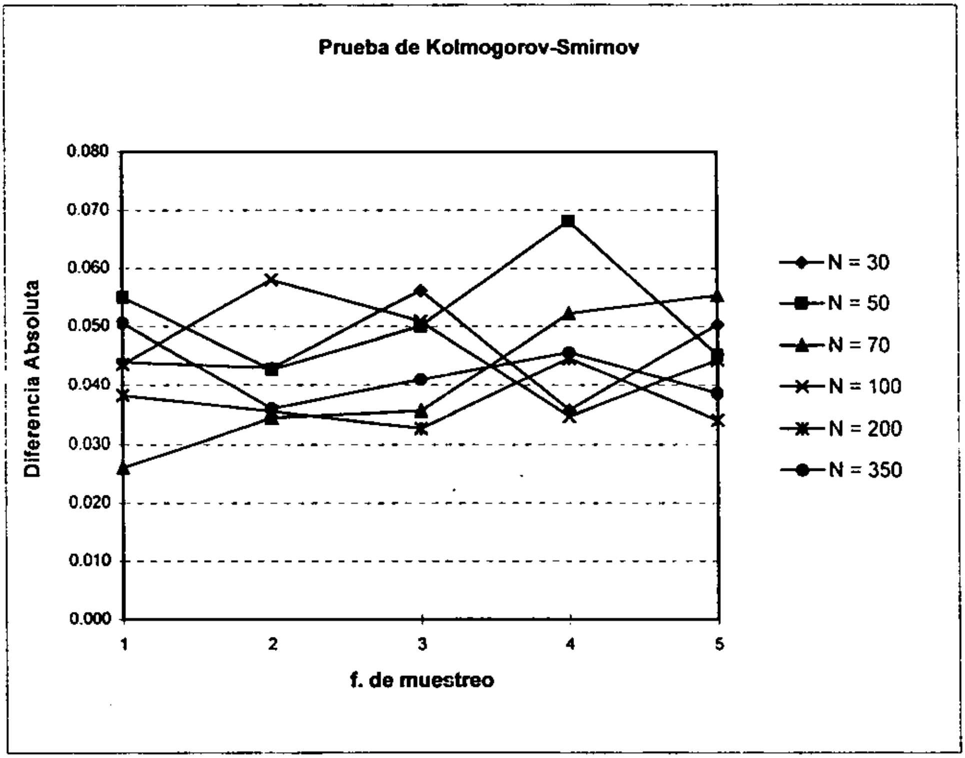
APÉNDICE B.2

EXPLICACIÓN A LOS:
CUADROS DE ESTADÍSTICAS CALCULADAS PARA LA PRUEBA DE KOLMOGOROV-SMIRNOV POR MODELO.

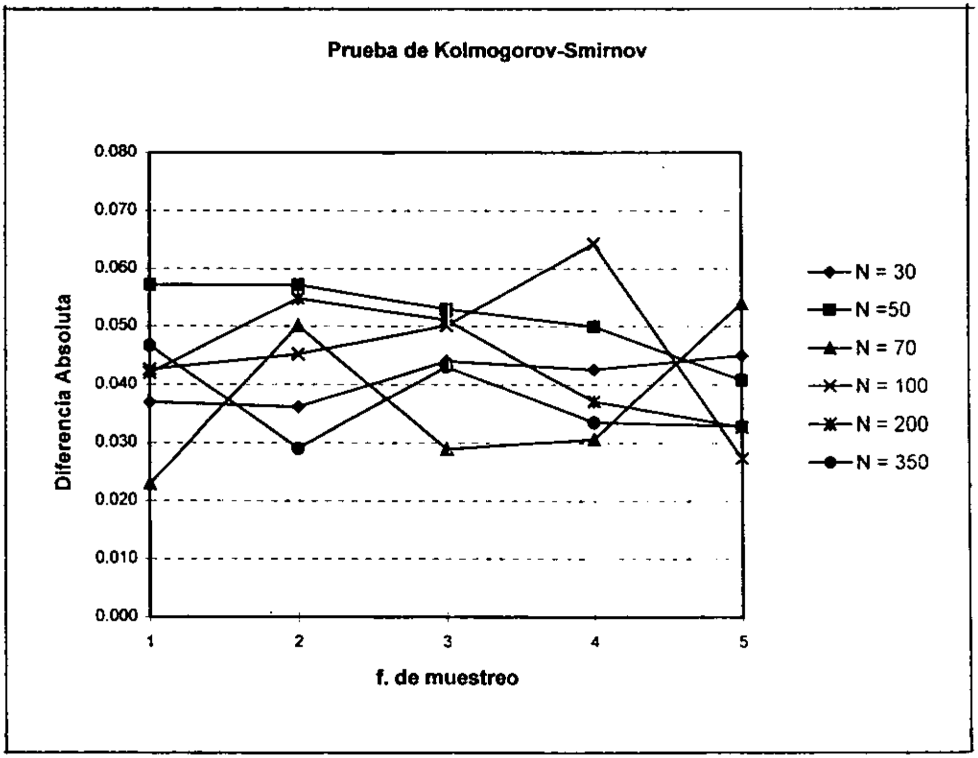
Bajo una correlación de la variable independiente con la variable dependiente y un cierto número total de elementos de la población y cinco porcentajes de muestreo.

Para cada uno de los quince casos anteriores, se efectuaron doscientas muestras, y a los resultados correspondientes para la estimación de la media en cada uno de los cinco procedimientos de estimación analizados, se aplicó la prueba de Kolmogorov-Smirnov. De las quince estadísticas calculadas para la prueba, se eligió la diferencia absoluta, cuyos valores se presentan en las tablas a continuación.

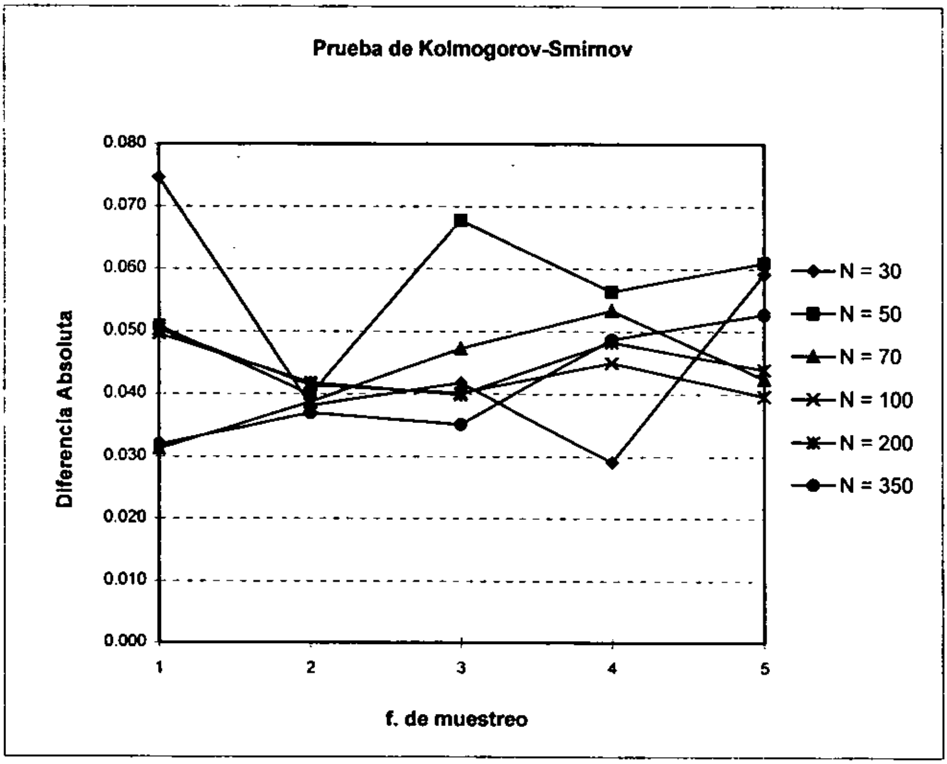
Prueba de Kolmogorov-Smirnov						
Simple			Cov. = 0.3			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.04395	0.05497	0.02594	0.04355	0.03821	0.05054
2	0.04295	0.04268	0.03447	0.05799	0.03555	0.03599
3	0.05612	0.05003	0.03564	0.05084	0.03270	0.04093
4	0.03576	0.06808	0.05211	0.03470	0.04455	0.04559
5	0.05029	0.04514	0.05528	0.04427	0.03411	0.03860



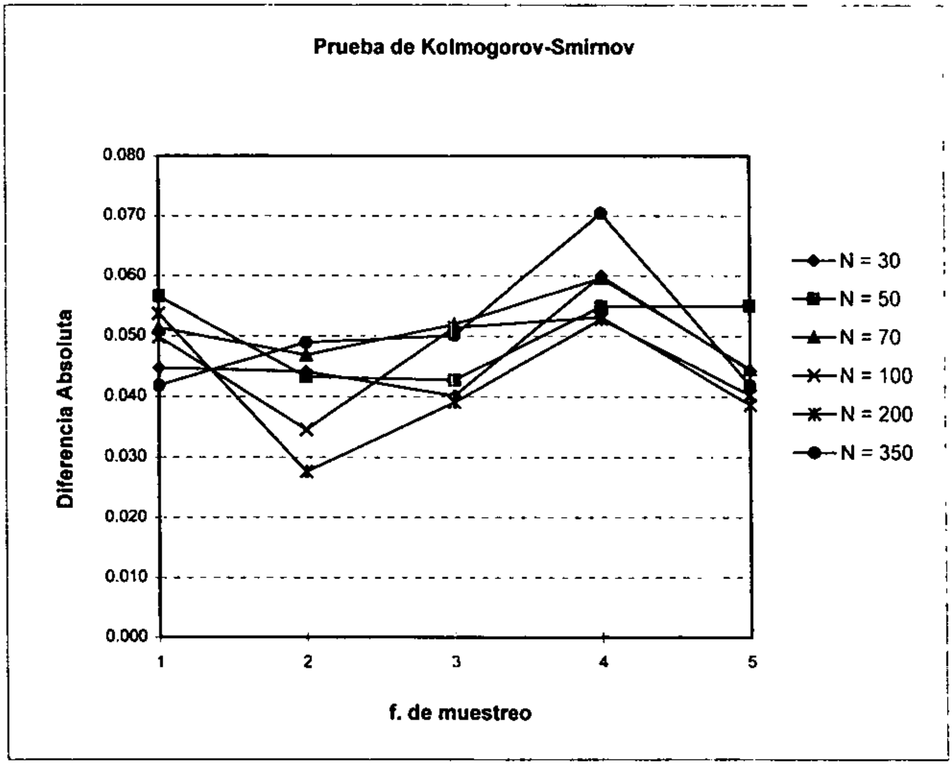
Prueba de Kolmogorov-Smirnov						
Razón			Cov. = 0.3			
f. de muestreo	N = 30	N =50	N = 70	N = 100	N = 200	N = 350
1	0.03691	0.05725	0.02294	0.04262	0.04199	0.04667
2	0.03606	0.05720	0.05010	0.04525	0.05481	0.02907
3	0.04408	0.05296	0.02895	0.05007	0.05105	0.04304
4	0.04261	0.04997	0.03061	0.06425	0.03704	0.03349
5	0.04508	0.04080	0.05398	0.02736	0.03272	0.03290



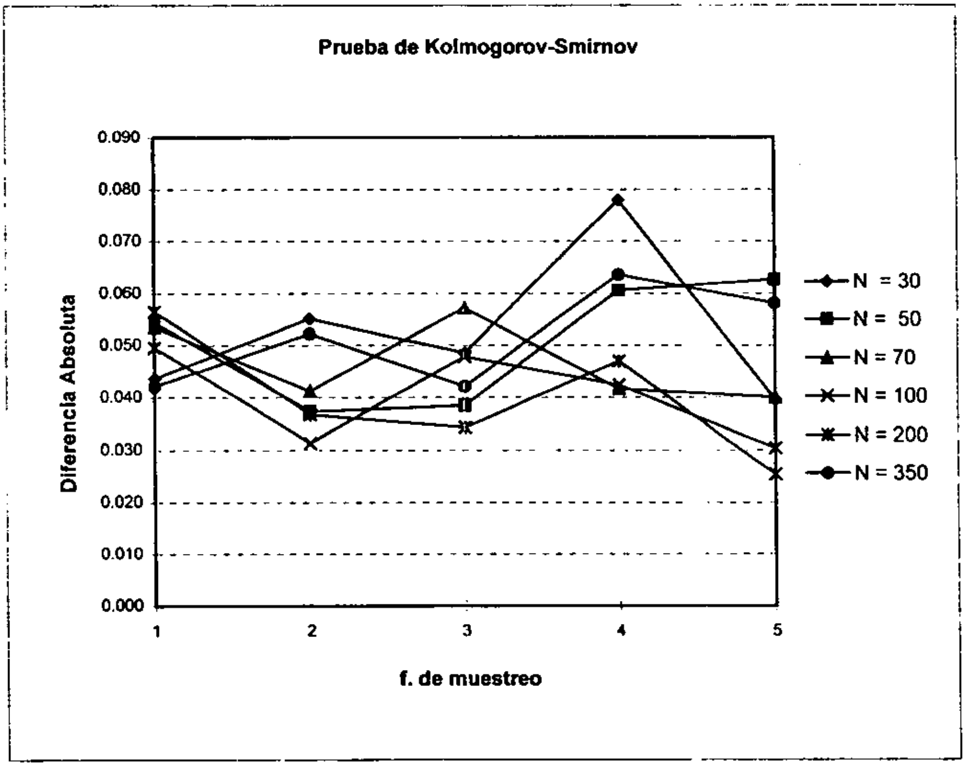
Prueba de Kolmogorov-Smirnov						
Regresión			Cov. = 0.3			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.07465	0.05084	0.03125	0.05023	0.04955	0.03194
2	0.03812	0.03981	0.03868	0.04133	0.04174	0.03695
3	0.04174	0.06776	0.04724	0.04009	0.03985	0.03516
4	0.02912	0.05632	0.05331	0.04493	0.04827	0.04869
5	0.05917	0.06095	0.04238	0.03961	0.04385	0.05272



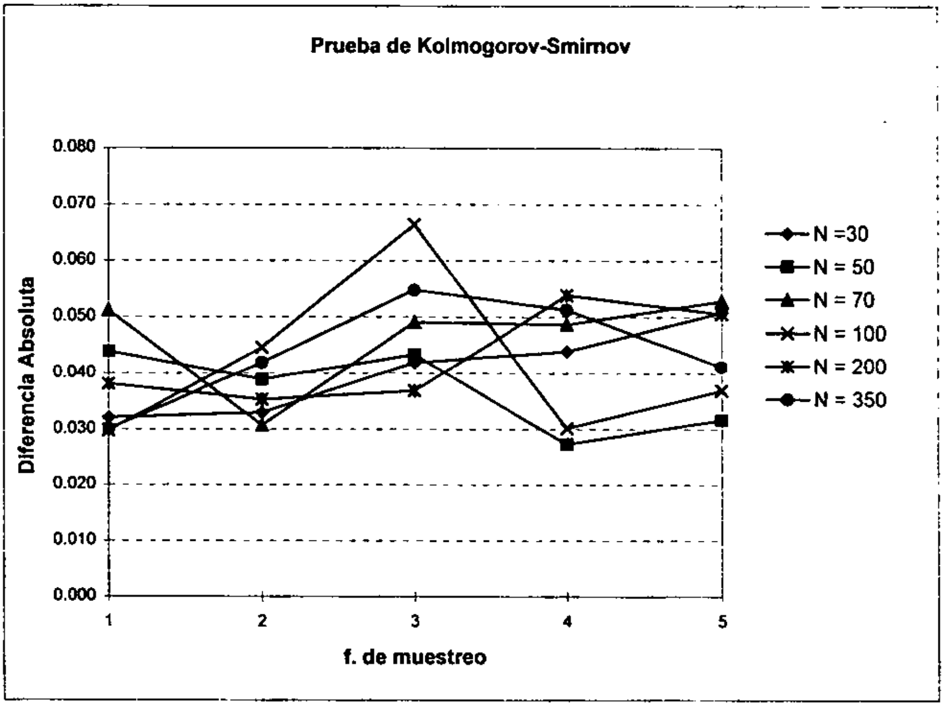
Prueba de Kolmogorov-Smirnov						
Secuencial			Cov. = 0.3			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.04469	0.05658	0.05137	0.04986	0.05371	0.0418
2	0.04404	0.04321	0.04682	0.03449	0.02756	0.04891
3	0.04009	0.04268	0.05191	0.05141	0.03901	0.05012
4	0.05998	0.05487	0.05959	0.05313	0.05281	0.07043
5	0.0443	0.055	0.04452	0.03857	0.04022	0.04184



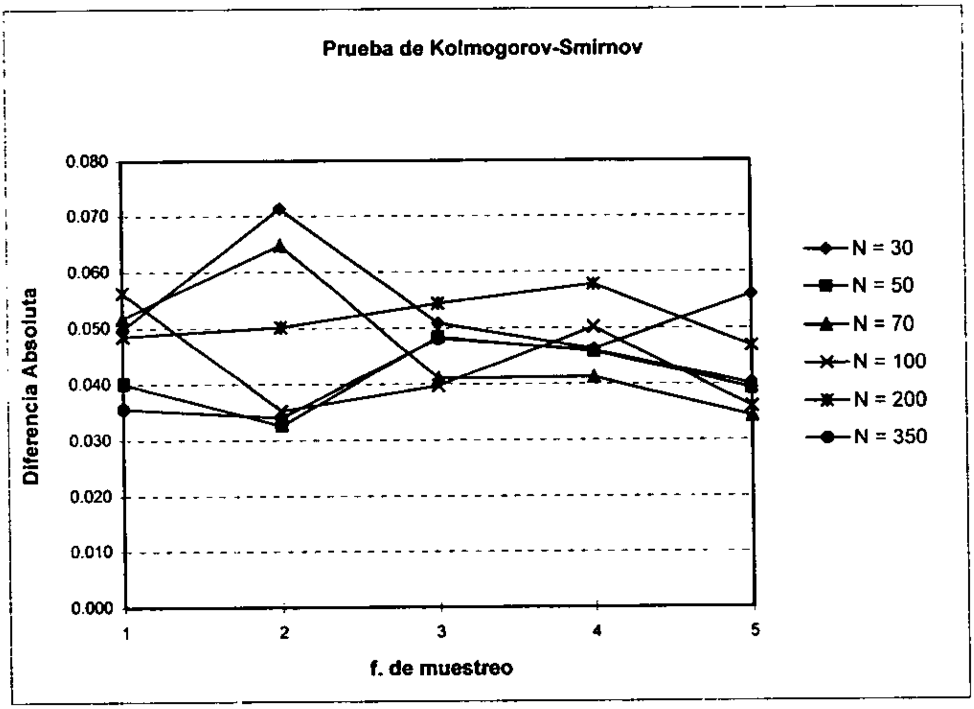
Prueba de Kolmogorov-Smirnov						
PPTWR			Cov. = 0.3			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.04365	0.05435	0.05341	0.04957	0.05647	0.04213
2	0.05511	0.03735	0.04122	0.03139	0.03684	0.0522
3	0.04851	0.03849	0.05714	0.04783	0.03431	0.04209
4	0.07795	0.06056	0.04162	0.04236	0.04698	0.06361
5	0.03954	0.06269	0.04002	0.03041	0.0253	0.05808



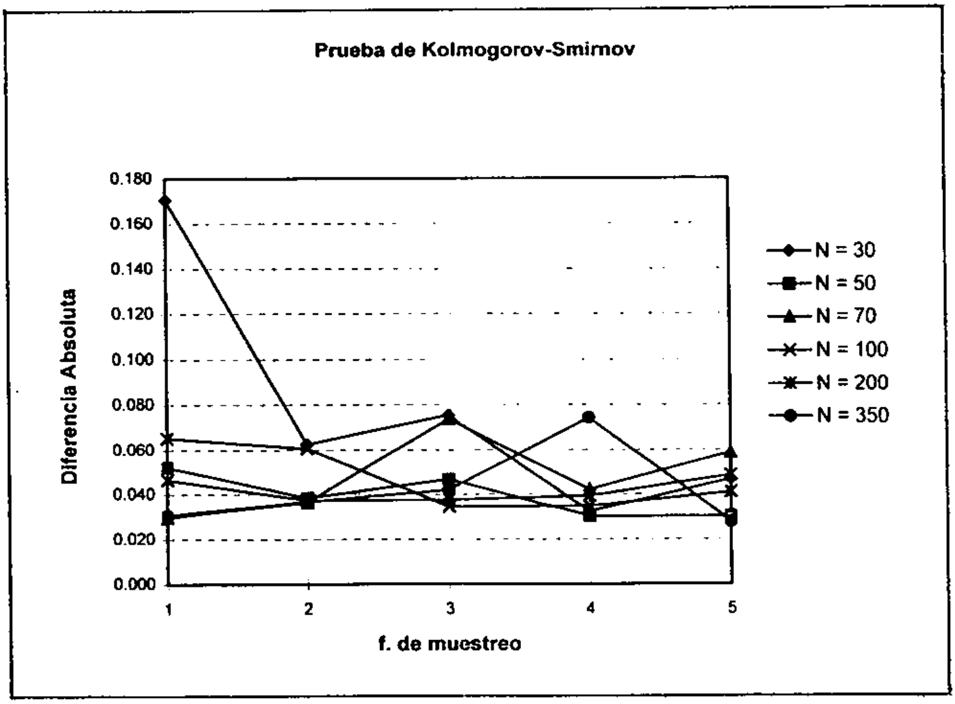
Prueba de Kolmogorov-Smirnov						
Simple			Cov. = 0.5			
f. de muestreo	N =30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.03208	0.04387	0.05113	0.02977	0.03804	0.02999
2	0.03295	0.03891	0.03072	0.04460	0.03532	0.04189
3	0.04189	0.04330	0.04907	0.06634	0.03686	0.05474
4	0.04391	0.02732	0.04866	0.03014	0.05377	0.05124
5	0.05079	0.03165	0.05278	0.03696	0.05056	0.04120



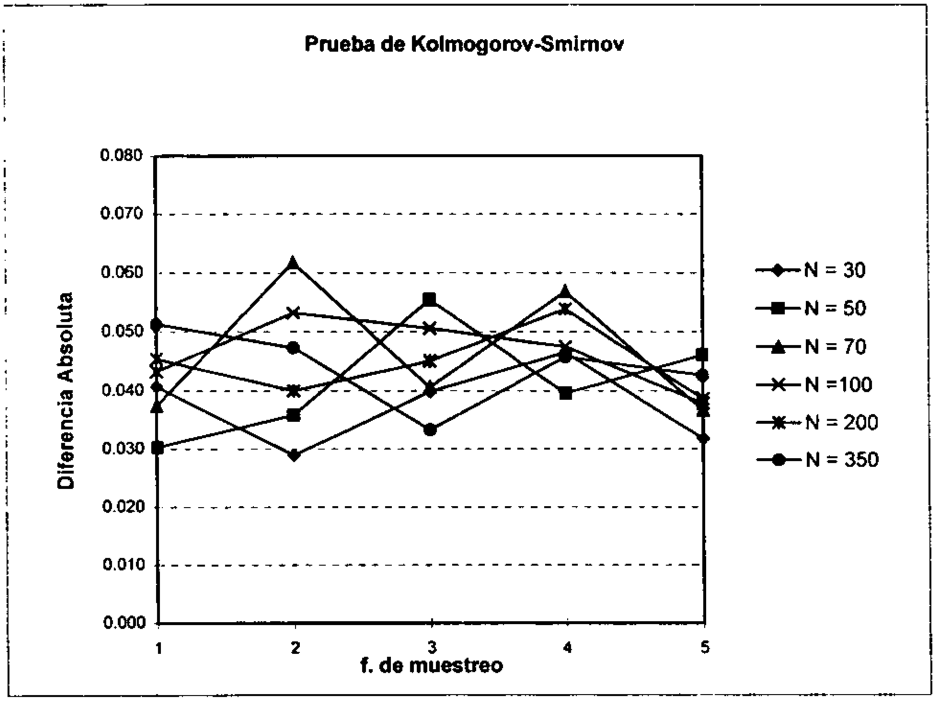
Prueba de Kolmogorov-Smirnov						
	Razón			Cov. = 0.5		
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.04981	0.04008	0.05175	0.05619	0.04859	0.03561
2	0.07132	0.03262	0.06483	0.03527	0.05014	0.03401
3	0.05083	0.04843	0.04097	0.03972	0.05435	0.04820
4	0.04634	0.04585	0.04115	0.05017	0.05770	0.04604
5	0.05587	0.03920	0.03429	0.03600	0.04682	0.03994



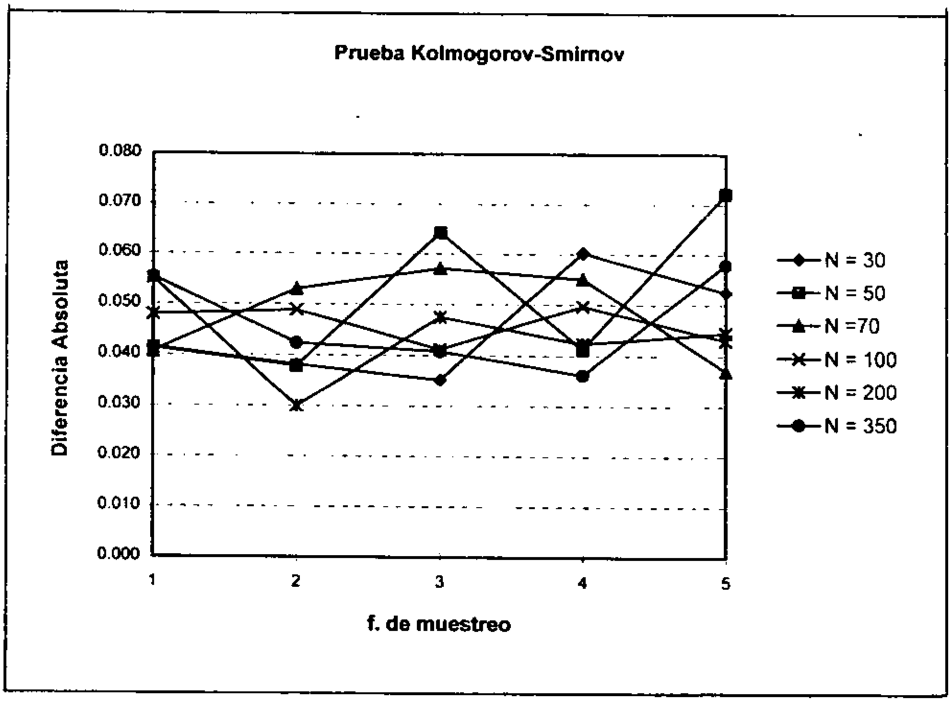
Prueba de Kolmogorov-Smirnov						
Regresión			Cov. = 0.5			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.17066	0.05214	0.02992	0.06529	0.04642	0.03082
2	0.06253	0.03837	0.03644	0.06066	0.03734	0.03681
3	0.07537	0.04661	0.07373	0.03446	0.03756	0.04190
4	0.03229	0.03029	0.04203	0.03474	0.03931	0.07433
5	0.04668	0.03018	0.05877	0.04094	0.04847	0.02774



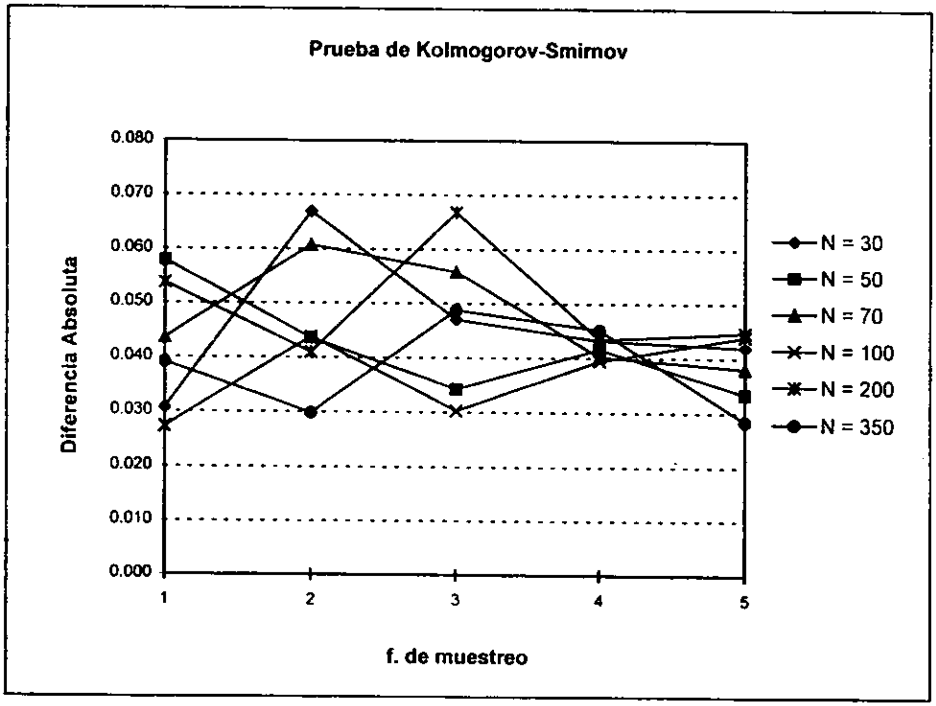
Prueba de Kolmogorov-Smirnov						
PPTWR			Cov. = 0.5			
f. de muestreo	N = 30	N = 50	N = 70	N =100	N = 200	N = 350
1	0.0407	0.03028	0.03741	0.04329	0.04543	0.05124
2	0.02893	0.03586	0.06169	0.05314	0.04	0.04733
3	0.03983	0.05544	0.04058	0.05057	0.04508	0.03333
4	0.04653	0.03956	0.05682	0.04749	0.05382	0.04582
5	0.0318	0.04614	0.03672	0.03783	0.03865	0.04264



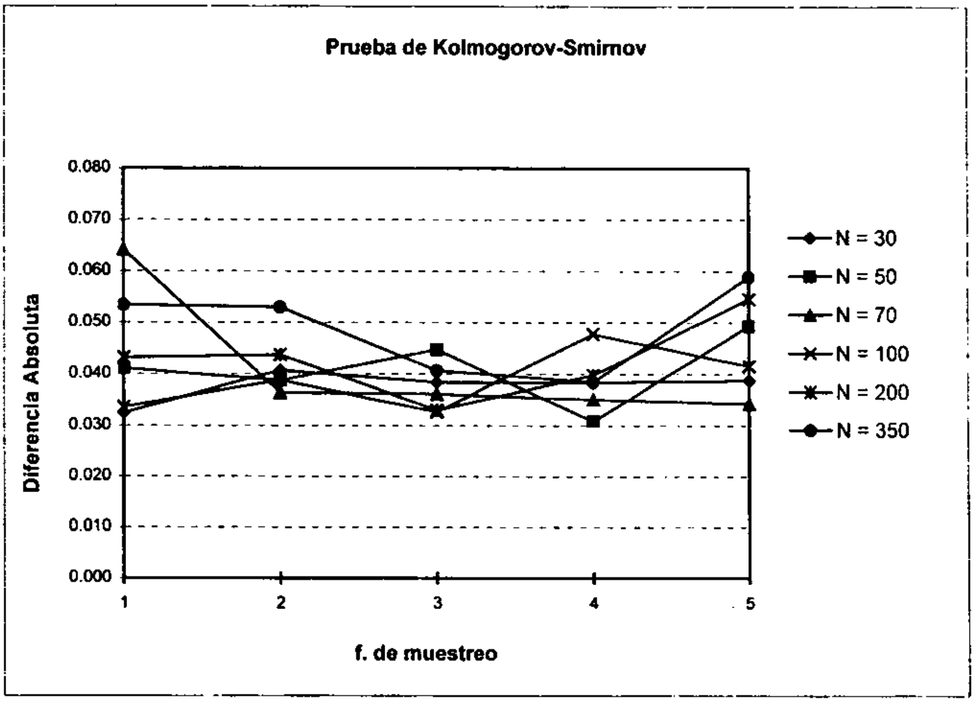
Prueba de Kolmogorov-Smirnov						
Secuencial				Cov. = 0.5		
f. de muestreo	N = 30	N = 50	N =70	N = 100	N = 200	N = 350
1	0.04151	0.04131	0.04062	0.04793	0.05504	0.05534
2	0.03811	0.0378	0.05305	0.04879	0.02992	0.04226
3	0.03522	0.06425	0.05713	0.04112	0.04745	0.04078
4	0.06025	0.04117	0.0551	0.04966	0.04227	0.03623
5	0.05247	0.07217	0.03704	0.04299	0.04444	0.05794



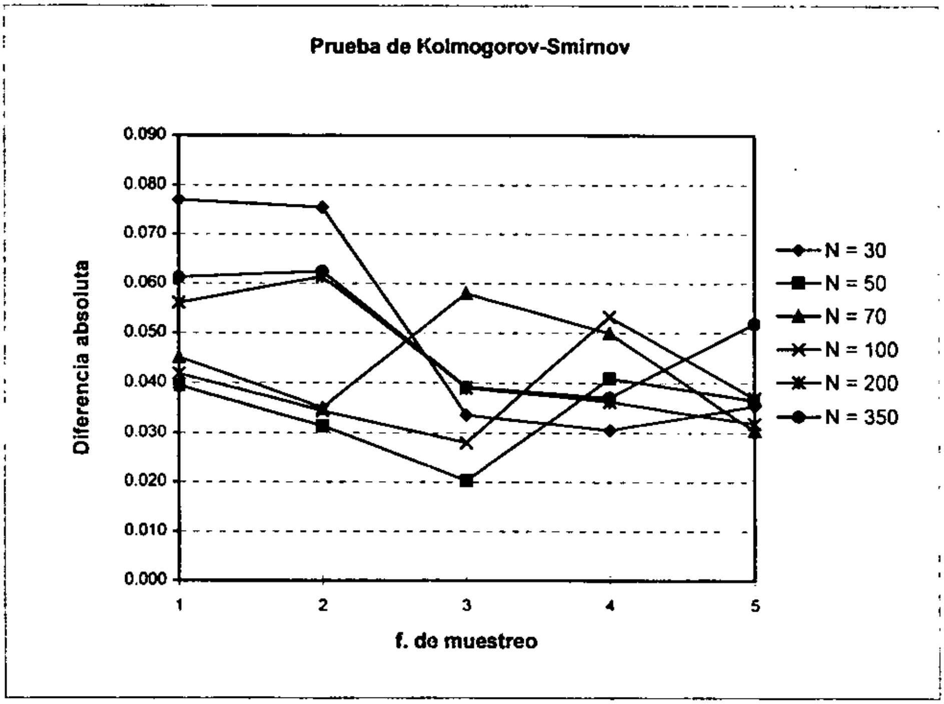
Prueba de Kolmogorov-Smirnov						
Simple			Cov. = 0.8			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.03064	0.05784	0.04356	0.02727	0.05369	0.03908
2	0.06691	0.04356	0.06072	0.04377	0.04095	0.02981
3	0.04699	0.03422	0.05579	0.03026	0.06677	0.04875
4	0.04324	0.04162	0.04023	0.03942	0.04340	0.04517
5	0.04195	0.03336	0.03803	0.04393	0.04466	0.02847



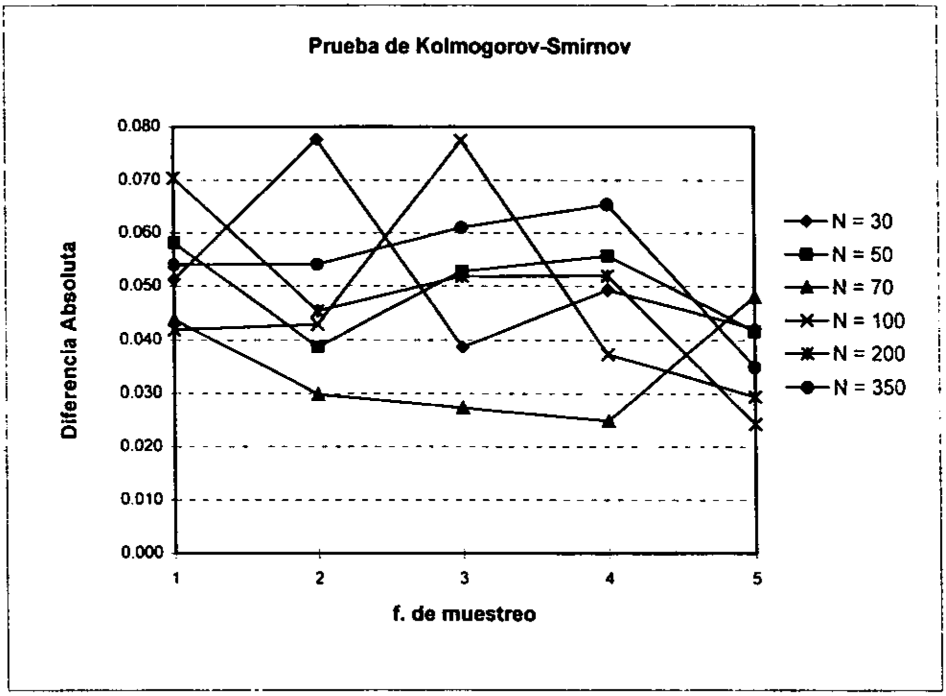
Prueba de Kolmogorov-Smirnov						
	Razón			Cov. = 0.8		
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.03248	0.04111	0.06412	0.03362	0.04321	0.05336
2	0.04072	0.03870	0.03635	0.03873	0.04376	0.05294
3	0.03836	0.04479	0.03616	0.03267	0.03302	0.04069
4	0.03837	0.03089	0.03511	0.04777	0.03982	0.03845
5	0.03873	0.04936	0.03431	0.04157	0.05455	0.05883



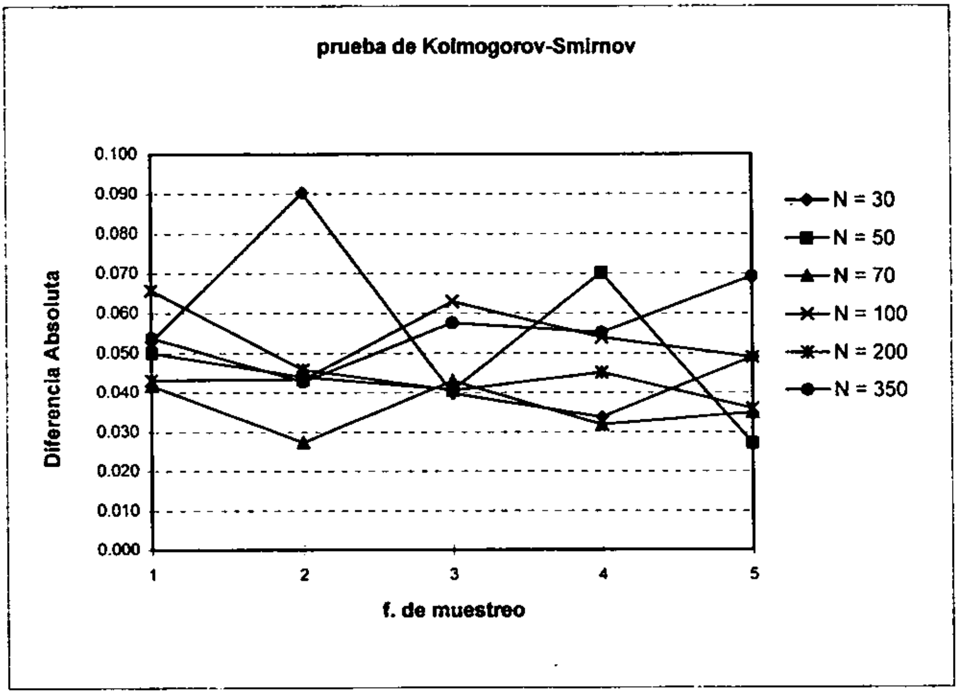
Prueba de Kolmogorov-Smirnov						
Regresión			Cov. = 0.8			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1	0.07697	0.03940	0.04514	0.04176	0.05609	0.06131
2	0.07542	0.03141	0.03497	0.03449	0.06143	0.06250
3	0.03369	0.02026	0.05802	0.02798	0.03901	0.03916
4	0.03057	0.04083	0.04997	0.05328	0.03632	0.03703
5	0.03548	0.03657	0.03045	0.03692	0.03191	0.05186



Prueba de Kolmogorov-Smirnov						
PPTWR			Cov. = 0.8			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1.00000	0.05122	0.05819	0.04372	0.04182	0.07039	0.05402
2.00000	0.07763	0.03863	0.02982	0.04284	0.04531	0.05412
3.00000	0.03862	0.05276	0.02739	0.07739	0.05178	0.06098
4.00000	0.04927	0.05574	0.02492	0.03720	0.05193	0.06535
5.00000	0.04213	0.04163	0.04805	0.02949	0.02433	0.03488



Prueba de Kolmogorov-Smirnov						
Secuencial			Cov. = 0.8			
f. de muestreo	N = 30	N = 50	N = 70	N = 100	N = 200	N = 350
1.00000	0.05290	0.05005	0.04175	0.04316	0.06571	0.05383
2.00000	0.09023	0.04397	0.02736	0.04333	0.04585	0.04272
3.00000	0.03970	0.04071	0.04294	0.06287	0.04048	0.05759
4.00000	0.03379	0.07016	0.03183	0.05393	0.04515	0.05516
5.00000	0.04903	0.02707	0.03490	0.04900	0.03593	0.06925



APÉNDICE C.1

PROGRAMA NOPPT

Programa que simuló el muestreo aleatorio simple sin reemplazo, desarrollando los estimadores directo, de razón y de regresión, este programa se desarrollo en FORTRAN 77, se tomó como base el programa desarrollado por Flores, así que sólo a continuación se muestra el programa principal y las subrutinas "models", "resul1", "archada", "vaver" y "massr" que fueron las que tuvieron modificaciones para la simulación de los modelos estudiados, las subrutinas "gepobg", "transf", "ajusta" fueron canceladas para la simulación del presente trabajo.

```
C*-----------------------------------------------------------------
C*
C*                  PROGRAMA
C*
C*         DESARROLLADO POR JOSE ANTONIO FLORES DIAZ
C*
C*           FECHA DE LA ULTIMA MODIFICACION
C*
C*         REALIZADA POR NANCY BEATRIZ  ZUÑIGA HIDALGO
C*
C*              8 DE SEPTIEMBRE DE 1996
C*-----------------------------------------------------------------
C*
C*           P R O G R A M A   P R I N C I P A L
C*
C*
C*-----------------------------------------------------------------
C*
C*  ESTE PROGRAMA DESARROLLA LA SIMULACION DEL  MUESTREO
C*
C*  ALEATORIO SIMPLE ESTIMANDO EL PARAMETRO DE LA MEDIA POR  TRES
C*
C*  ESQUEMAS COMO SON EL SIMPLE O DIRECTO, DE RAZON Y DE   REGRESION.
C*
C*  LO YA REFERIDO SE APLICA A LOS MODELOS SIGUIENTES:
C*
C*            1.- COV(X,Y) = 0.3
C*
C*            2.- COV(X,Y) = 0.5
C*
C*            3.- COV(X,Y) = 0.8
C*
C*
C*   LOS ARREGLOS QUE UTILIZA SON:
```

```
C*
C*  ARREGLO XPOB                    ESTE ARREGLO CONTENDRA LOS VALORES
C*  REAL                            DE LAS 'X'
C*
C*  ARREGLO NELEM                   ESTE ARREGLO CONTENDRA LOS VALORES
C*  ENTERO                          DE LOS TAMANIOS DE POBLACION QUE SE
C*                                  ESTAN CONSIDERANDO
C*
C*   ARREGLO YPOB                   ESTE ARREGLO CONTENDRA LOS VALORES
C*   REAL                           DE LOS DIFERENTES MODELOS CONSIDERADOS
C*
C*   ARREGLO RMUPOB                 ESTE ARREGLO CONTENDRA LOS VALORES
C*   REAL                           CORRESPONDIENTES A LAS MEDIAS PO-
C*                                  BLACIONALES DE LOS DIFERENTES MODELOS
C*                                  CONSIDERADOS
C*
C*   ARREGLO MCOV                   ESTE ARREGLO CONTENDRA LOS VALORES DE LAS
C*   REAL                           COVARIANZAS CORRESPONDIENTES A CADA MODELO
C*
C*   ARREGLO VARPOB         ESTE ARREGLO CONTENDRA LOS VALORES
C*   REAL                   CORRESPONDIENTES A LAS VARIANZAS
C*                          POBLACIONALES DE LOS DIFERENTES MO-
C*                          DE LOS CONSIDERADOS
C*
C*
C*  FORMA DE UTILIZARSE  - NINGUNA PORQUE SE TRATA DEL PROGRAMA PRINCI-
C*                 PAL
C*
C*  SUBRUTINAS QUE EMPLEA
C*
C*                - MODELS
C*
C*                - TOTPOB
C*
C*                - RESULI
```

```
C*
C*                - ARCHDA
C*
C*                - MASSR
C*
C*-----------------------------------------------------------------
C*
      IMPLICIT NONE
      DIMENSION XPO(500),NELEM(9),YPOB(500),RMUPOB(2),VARPOB(2),

     *XPOB(500),MCOV(3)

      REAL XPO,YPOB,RMUPOB,XPOB,SUMX,SUMX2,VARPOB,MCOV

      INTEGER NELEM,I,J,L,DSEED,DSEED1

C*                   PRIMERA INSTRUCCION EJECUTABLE

      DATA NELEM/30,50,70,100,200,350/

      DATA MCOV/0.3,0.5,0.8/

      OPEN (6,FILE="FINAL1.")
      OPEN(2,FILE="MEDIAS.")
      OPEN (1,FILE="DATOS.")
      OPEN (3,FILE="MEDSVARS.")
      OPEN (7,FILE="MONOPPT1.")
      OPEN (8,FILE="MONOPPT2.")
      OPEN (9,FILE="MONOPPT3.")

       DSEED = 12357987

       DSEED1 = DSEED
       DO 2 I = 1 , 6
```

```
      DO 4 L = 1,3
        CALL MODELS(XPOB,YPOB,NELEM(I),MCOV(L),SUMX,SUMX2)

        CALL TOTPOB(YPOB,NELEM(I),RMUPOB,VARPOB,SUMX,SUMX2)

        CALL RESUL1(RMUPOB,VARPOB,NELEM(I),MCOV(L))

        CALL ARCHDA(XPOB,YPOB,NELEM(I))

        CALL MASSR(XPOB,YPOB,NELEM(I),RMUPOB,VARPOB,DSEED1,L)
  4     CONTINUE
  2  CONTINUE
     CLOSE(6)

     CLOSE(2)

     CLOSE(1)

     CLOSE(3)

     CLOSE(7)

     CLOSE(8)

     CLOSE(9)

     CALL EXIT

     END
C*
C* SUBRUTINA MODELS
C*
C*---------------------------------------------------------------------
C*
C* ESTA SUBRUTINA CALCULA O EVALUA LOS DIFERENTES MODELOS QUE
```

```
C*
C* SE ESPECIFICARON YA EN EL PROGRAMA PRINCIPAL, ADEMAS EVALUA
C*
C* LA SUMA DE LAS X COMO TAMBIEN LA SUMA DE LAS X AL CUADRADO
C*
C*   LOS PARAMETROS DE ESTA SUBRUTINA SON LOS SIGUIENTES:
C*
C*    ARREGLO X            SALIDA CONTENDRA LOS VALORES DE LAS 'X'
C*    REAL
C*
C*    ARREGLO Y            SALIDA CONTENDRA LOS VALORES DE LOS
C*    REAL                 DIFERENTES  MODELOS
C*
C*    VARIABLE  NR         ENTRADA CONTENDRA EL TAMANIO DE LA POBLACION
C*    ENTERA       O NUMERO DE ELEMENTOS DE LA POBLACION
C*
C*   VARIABLE MCOV         ENTRADA CONTENDRA EL VALOR DE LA COVARIANZA
C*   REAL                  DE LOS DIFERENTES MODELOS
C*
C*    VARIABLE SUMX        SALIDA CONTENDRA EL VALOR CORRESPONDIENTE A
C*    REAL                 LA SUMA DE LAS X
C*
C*    VARIABLE SUMX2       SALIDA CONTENDRA EL VALOR CORRESPONDIENTE A
C*    REAL                 LA  SUMA DE LAS X AL CUADRADO
C*
C*   FORMA DE USO  - CALL MODELS(X,Y,N,MCOV,SUMX,SUMX2)
C*
C*   SUBRUTINAS QUE EMPLEA  -
C*                 CHFAC  (IMSL)
C*                 RNMUV  (IMSL)
C*                 RNSET  (IMSL)
C*                 UMACH  (IMSL)
C*
C*-----------------------------------------------------------------
C*
```

```
      SUBROUTINE MODELS(X,Y,NR,MCOV,SUMX,SUMX2)

C*                      ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE

      INTEGER NR,IRANK,K,LDR,LDRSIG,NOUT,LI,I,ISEED
      REAL COV(2,2), R(500,2),RSIG(2,2)

      REAL SUMX,SUMX2,X2,X,MCOV,Y
C*
      EXTERNAL CHFAC,RNMVN,RNSET,UMACH
C*
      DIMENSION X(NR), Y(500)

      CALL UMACH(2,NOUT)

      ISEED = 12357987

      LDR = 500

      K = 2

      LDRSIG = 2

      COV(1,1)=1.0

      COV(1,2)=MCOV

      COV(2,1)=COV(1,2)

      COV(2,2)=COV(1,1)

C*     OBTIENE LA FACTORIZACION DE CHOLESKY
```

```
      CALL CHFAC(K,COV,2,0.00001,IRANK,RSIG,LDRSIG)

      CALL RNSET(ISEED)

      CALL RNMVN(NR,K,RSIG,LDRSIG,R,LDR)

      SUMX=0.0

      SUMX2=0.0

      DO 12 I= 1,NR

C*    SE INCREMENTA UNA CONSTANTE C=10 A LOS DATOS PARA EVITAR

C*    VALORES NEGATIVOS EN LAS VARIABLES `X`

      X(I)= R(I,1)+10

C*    CALCULA X AL CUADRADO

      X2 = X(I)*X(I)

C*    CALCULA LA SUMA DE LAS X

      SUMX=SUMX+X(I)

C*    CALCULA LA SUMA DE LAS X AL CUADRADO

      SUMX2= SUMX2 + X2

C*   DETERMINA EL Y(I) CORRESPONDIENTE
C*
      Y(I)= R(I,2)

   12 CONTINUE
```

```
      RETURN

      END
C*
C*SUBRUTINA RESUL1
C*------------------------------------------------------------------------------------------------
C*
C* ESTA SUBRUTINA SE ENCARGA DE IMPRIMIR LOS RESULTADOS CORRESPON-
C*
C* DIENTES A LA MEDIA Y A LA VARIANZA POBLACIONALES DE LAS VARIA-
C*
C* BLES AUXILIAR Y DEL RESTO DE LOS MODELOS Y'ES CONSIDERADOS
C*
C* LOS PARAMETROS QUE EMPLEA SON :
C*
C*
C*    ARREGLO RMUPOB          ENTRADA CONTENDRA LA INFORMACION DE ME-
C*    REAL                    DIAS PARA LA POBLACION TANTO DE SU CARAC-
C*                            TERISTICA AUXILIAR COMO LA DE LOS MODELOS
C*                            QUE SE ESTAN CONSIDERANDO
C*
C*    ARREGLO VARPOB          ENTRADA CONTENDRA LA INFORMACION DE VA-
C*    REAL                    RIANZAS PARA LA POBLACION TANTO DE SU
C*                            CARACTERISTICA AUXILIAR COMO LA DE LOS
C*                            MODELOS QUE SE ESTAN CONSIDERANDO
C*
C*    VARIABLE MCOV           ENTRADA CONTENDRA EL VALOR
C*    REAL                    CORRESPONDIENTE DE LA COVARIAZA EN
C*                            ESTUDIO
C*
C*    VARIABLE N              ENTRADA CONTENDRA LA INFORMACION DE EL
C*    ENTERA                  NUMERO DE ELEMENTOS QUE COMPONEN LA PO-
C*                            BLACION
C*
```

```
C*
C* LA FORMA DE USO   -CALL RESUL1(RMUPOB,VARPOB,N,MCOV)
C*
C* SUBRUTINAS QUE EMPLEA  - NINGUNA
C*
C*-------------------------------------------------------------
C*
      SUBROUTINE RESUL1(RMUPOB,VARPOB,N,MCOV)

C*                 ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE

      DIMENSION RMUPOB(2),VARPOB(2),TIT(2)

      REAL RMUPOB,VARPOB,TIT,RMUS,MCOV

      INTEGER N


C*                 ESPECIFICACION DE ELEMENTOS LOCALES

      INTEGER J,K

C*                 PRIMERA INSTRUCCION EJECUTABLE

      TIT(1)= "X"

      TIT(2)= "Y"

      WRITE(6,200) N, MCOV

      WRITE(6,201)

      DO 1 K = 1 , 2
```

```
      WRITE(6,202)TIT(K),RMUPOB(K),VARPOB(K)
 1    CONTINUE
      WRITE(6,203)
      RETURN

200   FORMAT(1H1,11X,"T A B L A  D E  V A L O R E S  P O B L A",
     *"C I O N A L E S",//,36X,"( N = ",I3," )",//,36X,"(COV=",F5.3,")",
     *//,10X,50("-"))
201   FORMAT(10X,"I MODELO O CONCEPTO  I",4X,"MEDIA",4X,"I",2X,"VA",
     *"RIANZA",3X,"I")
202   FORMAT(10X,"I ",9X,A1,8X," I ",F11.5," I ",F11.5," I")
203   FORMAT(10X,50("-"))
      END

C* SUBRUTINA ARCHDA
C*
C*-----------------------------------------------------------------
C*
C* ESTA SUBRUTINA SE ENCARGA DE GENERAR EL ARCHIVO DONDE SE ESCRI-
C*
```

```
C* BEN LOS DATOS CON LOS QUE POSTERIORMENTE TRABAJAN LAS DEMAS SUB-
C*
C* RUTINAS DE ESTE PROGRAMA. EL PROPOSITO ES GENERAR UN ARCHIVO EL
C*
C* CUAL PUEDA SER ANALIZADO POSTERIORMENTE POR EL SPSS
C*
C* LOS ARGUMENTOS DE ESTA SUBRUTINA SON :
C*

C*     ARREGLO X              ENTRADA CONTENDRA EL CONJUNTO DE DATOS
C*     REAL                   QUE SIRVEN DE BASE PARA LA GENERACION DE
C*                            LOS MODELOS
C*
C*     ARREGLO Y              ENTRADA CONTENDRA EL CONJUNTO DE DATOS
C*     REAL                   CORRESPONDIENTES A LOS MODELOS YA GENE-
C*                            RADOS
C*
C*     VARIABLE N             ENTRADA CONTENDRA LA INFORMACION CO-
C*     ENTERA                 RRESPONDIENTE AL NUMERO DE UNIDADES
C*                            QUE COMPONEN A LA POBLACION
C*
C*  - FORMA DE USO    - CALL ARCHDA(X,Y,N)
C*
C*  - SUBRUTINAS QUE EMPLEA   - NINGUNA
C*
C*------------------------------------------------------------------
C*
      SUBROUTINE ARCHDA(X,Y,N)

C*                   ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE


      INTEGER N
```

```
      DIMENSION X(N),Y(500)

      REAL X,Y

C*                   ESPECIFICACION DE ELEMENTOS LOCALES

      INTEGER I

C*                   PRIMERA INSTRUCCION EJECUTABLE

        DO  1  I = 1 , N

          WRITE(1,*) X(I),Y(I)

    1   CONTINUE

        RETURN

        END

C*
C* SUBRUTINA MASSR
C*
C*-----------------------------------------------------------------------------------------------------------------
C*
C* ESTA SUBRUTINA SE ENCARGA DEL PRINCIPAL PROCESO EN LA SIMULACION
C*
C* DADO QUE EN ELLA SE REALIZA TODO LO REFERENTE A SELECCIONAR UNA
C*
C* MUESTRA ALEATORIA SIMPLE, OBTENER RESULTADOS PARA EL PARAMETRO DE
C*
C*LA  MEDIA BAJO LOS ESQUEMAS DE CALCULO DIRECTO, DE RAZON Y DE REGRE-
C*
C* SION.   FINALMENTE CUANDO SE HA LLEGADO AL NUMERO DE MUESTRAS QUE
```

```
C*
C*SE DESEAN OBTENER ENTONCES SE PROSIGUE CON OTRO PORCENTAJE DE
C*
C*MUESTREO  O SE REGRESA AL PROGRAMA PRINCIPAL
C*
C* TODO LO ANTERIOR LO DESARROLLA PARA LOS PORCENTAJES DE
C*
C* MUESTREO SIGUIENTES : 10, 20, 30, 40 Y 45
C*
C* LOS PARAMETROS QUE EMPLEA SON :
C*
C*    ARREGLO X           ENTRADA CONTENDRA LA INFORMACION DE LA
C*    REAL                 POBLACION AUXILIAR
C*
C*    ARREGLO Y           ENTRADA CONTENDRA LA INFORMACION DE LOS
C*    REAL                DIFERENTES MODELOS PARA LA POBLACION EN
C*                        ESTUDIO
C*
C*    VARIABLE N          ENTRADA CONTENDRA LA INFORMACION DE EL
C*    ENTERA              NUMERO DE ELEMENTOS QUE COMPONEN LA PO-
C*                        BLACION
C*
C*    ARREGLO RMUPOB      ENTRADA CONTENDRA LA INFORMACION DE ME-
C*    REAL                DIAS PARA LA POBLACION TANTO DE SU CARAC-
C*                        TERISTICA AUXILIAR COMO LA DE LOS MODELOS
C*                        QUE SE ESTAN CONSIDERANDO
C*
C*    ARREGLO VARPOB      ENTRADA CONTENDRA LA INFORMACION DE VA-
C*    REAL                 RIANZAS PARA LA POBLACION TANTO DE SU
C*                        CARACTERISTICA AUXILIAR COMO LA DE LOS MO-
C*                        DE LOS QUE SE ESTAN CONSIDERANDO
C*
C*    VARIABLE DSEEMA     ENTRADA/SALIDA CONTENDRA LA INFORMACION
C*    ENTERA              CORRESPONDIENTE PARA QUE LAS SUBRUTINAS
C*                        DEL IMSL SE USEN
```

```
C*
C*
C*    VARIABLE L          ENTRADA CONTENDRA LA INFORMACION QUE
C*    ENTERA              DETERMINA QUE MODELO SE ESTA CONSIDERANDO
C*
C*      ++++++++++++++++++++++++++++++++++++++++
C*
C* OTROS ARREGLOS O ELEMENTOS QUE SE USAN EN LA SUBRUTINA SON
C*
C*    ARREGLO PORMUE      SENIALA LOS PORCENTAJES DE MUESTREO
C*    REAL
C*
C*    ARREGLO IX          CONTENDRA LA INFORMACION DE LOS
C*    ENTERO              INDICES DE LA POBLACION QUE SON ELEGIDOS
C*                        EN UNA MUESTRA
C*
C*    ARREGLO REMUEM      CONTENDRA LA MEDIA ESTIMADA BAJO CADA ES-
C*    REAL                QUEMA PARA LOS DIFERENTES MODELOS QUE SE
C*                        ESTAN CONTEMPLANDO
C*
C*    ARREGLO REMUEV      CONTENDRA LA VARIANZA ESTIMADA BAJO CADA
C*    REAL                ESQUEMA PARA LOS DIFERENTES MODELOS QUE SE
C*                        ESTAN COMTEMPLANDO
C*
C*    ARREGLO VARVE       CONTENDRA LAS VARIANZAS VERDADERAS QUE
C*    REAL                CORRESPONDERIAN A LOS ESQUEMAS DE CALCU-
C*                        LO QUE SE DESARROLLARON
C*
C*    ARREGLO SVARVE      CONTENDRA LAS RAICES CUADRADAS DE LAS VA-
C*    REAL                RIANZAS VERDADERAS QUE CORRESPONDERIAN A
C*                        LOS ESQUEMAS DE CALCULO QUE SE DESARROLLA-
C*                        RON
C*
C*    ARREGLO TEMP        CONTENDRA LAS DIFERENCIAS ENTRE LAS MEDIAS
C*    REAL                ESTIMADAS Y LAS VERDADERAS PARA CADA MUES-
```

```
C*                           TRA REALIZADA
C*

C*   ARREGLO SXYP         CONTENDRA LA COVARIABILIDAD ENTRE LA VA-
C*.  REAL                 RIABLE EQUIS Y LA CORRESPONDIENTE VARIA-
C*                        BLE YE
C*
C*      ++++++++++++++++++++++++++++++++++++++++++++++++++++
C*
C* FORMA DE USO  - CALL MASSR(X,Y,N,RMUPOB,VARPOB,DSEEMA,L)
C* SUBRUTINAS QUE EMPLEA
C*
C*           - RNSET (IMSL)
C*
C*           - VAVER
C*
C*           - RNSRI (IMSL)
C*
C*           - SUMUES
C*
C*           - SXYMUE
C*
C*           - CALCU
C*
C*           - RNGET (IMSL)
C*
C*-----------------------------------------------------------------
C*
      SUBROUTINE MASSR(X,Y,N,RMUPOB,VARPOB,DSEEMA,L)

C*                    ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE
C*
      INTEGER N,IMPRI,DSEEMA,L
```

```
      DIMENSION X(N),Y(500),RMUPOB(2),VARPOB(2)
      REAL X,Y,RMUPOB,VARPOB
C*                    ESPECIFICACION DE ELEMENTOS LOCALES
      DIMENSION PORMUE(5),IX(500),REMUEM(3),
     *REMUEV(3),VARVE(3),SVARVE(3),TEMP(3)
      REAL PORMUE,REMUEM,TNMUE,REMUEV,F,F1,VARVE,SVARVE,SX,SX2,
     *SY,SY2,SXY,TEMP,SXYP
      INTEGER IX,NMUE,IENS,I,J,K,M
C*                    PRIMERA INSTRUCCION EJECUTABLE
      DATA PORMUE/.1,.2,.3,.4,.45/
      CALL RNSET(DSEEMA)
      DO 4 I = 1 , 5
        IENS = 0
        TNMUE = PORMUE( I ) * FLOAT( N )
        NMUE = AINT( TNMUE )
        IF( TNMUE - NMUE .GT. 0.00000001 ) NMUE = NMUE + 1
        F = FLOAT( NMUE ) / FLOAT( N )
```

```
      IF( L .EQ. 1 ) WRITE(2,200) N,NMUE
      F1 = ( 1.0 - F ) / ( FLOAT( NMUE ) * ( FLOAT( NMUE ) - 1.0 ))
1     CALL RNSRI(NMUE,N,IX)
      IENS = IENS + 1
      CALL SUMUES(X,N,IX,NMUE,SX)
      CALL SXYMUE(X,X,N,IX,NMUE,SX2)
        IF ( VARPOB(2) .EQ. 0.0 ) GO TO 3
        CALL SUMUES(Y,N,IX,NMUE,SY)
        CALL SXYMUE(Y,Y,N,IX,NMUE,SY2)
        CALL SXYMUE(Y,X,N,IX,NMUE,SXY)
        DO 2 K = 1 , 3
          CALL CALCU(K,SY,SX,SXY,SY2,SX2,NMUE,F1,RMUPOB(1),
     *          REMUEM(K),REMUEV(K))
          TEMP(K) = REMUEM(K) - RMUPOB(2)
2       CONTINUE
        K = L + 6
        WRITE(K,202) L,IENS,(REMUEM(M),M=1,3),(REMUEV(M),
     *       M=1,3)
```

```
  3 CONTINUE

    IF(IENS .LT. 200) GO TO 1

  4 CONTINUE


   CALL RNGET(DSEEMA)

   RETURN

200 FORMAT(2I3)


202 FORMAT(I1,I3,6F11.5)

   END

C* SUBRUTINA VAVER
C*-------------------------------------------------------------------------------------------------------
C* ESTA SUBRUTINA CALCULA LAS VARIANZAS MUESTRALES QUE CORRESPONDE-
C*
C* RIAN UTILIZANDO LOS VALORES POBLACIONALES CON EL FIN DE PODER REA-
C*
C* LIZAR COMPARACIONES ENTRE LAS ESTIMACIONES QUE SE HACEN CON BASE
C*
C* A TODAS Y CADA UNA DE LAS MUESTRAS Y LOS VALORES QUE AQUI SE CAL-
C*
C* CULAN
C*
C*  LOS PARAMETROS QUE SE EMPLEAN SON:
C*
C*    ARREGLO RMU              ENTRADA CONTENDRA LA INFORMACION DE LAS
C*    REAL                     MEDIAS POBLACIONALES DE LA VARIABLE AU-
```

```
C*                              XILIAR Y LOS 3 MODELOS CONSIDERADOS
C*
C*    ARREGLO VAR               ENTRADA CONTENDRA LA INFORMACION DE LAS
C*    REAL                      VARIANZAS POBLACIONALES DE LA VARIABLE
C*                              AUXILIAR Y LOS 3  MODELOS CONSIDERADOS
C*
C*    VARIABLE F                ENTRADA CONTENDRA LA INFORMACION
C*    REAL                      CORRESPONDIENTE A LA FRACCION DE
C*                              MUESTREO
C*
C*    VARIABLE NMUE             ENTRADA CONTENDRA LA INFORMACION
C*    ENTERA                    CORRESPONDIENTE AL TAMANIO DE MUESTRA
C*
C*    ARREGLO Y                 ENTRADA CONTENDRA LA INFORMACION
C*    REAL                      CORRESPONDIENTE A LOS MODELOS
C*                              CONSIDERADOS
C*
C*    ARREGLO X                 ENTRADA CONTENDRA LA INFORMACION
C*    REAL                      CORRESPONDIENTE A LA VARIABLE AUXILIAR
C*
C*    VARIABLE N                ENTRADA CONTENDRA LA INFORMACION
C*    ENTERA                    CORRESPONDIENTE A EL TAMANIO DE LA
C*                              POBLACION
C*
C*    ARREGLO VAR1              SALIDA CONTENDRA LA INFORMACION CORRES-
C*    REAL                      PONDIENTE A LOS VALORES DE LAS VARIANZAS
C*                              QUE CORRESPONDERIAN A CADA ESQUEMA DE
C*                              CALCULO
C*
C*    ARREGLO SVAR1             SALIDA CONTENDRA LAS RAICES CUADRADAS
C*    REAL                      DE LAS VARIANZAS VERDADERAS QUE
C*                              CORRESPONDERIAN A    CADA ESQUEMA DE
C*                              CALCULO
C*
C*
```

```
C*    ARREGLO SXY                 ENTRADA/SALIDA CONTENDRA LA
C*    REAL                        COVARIABILIDAD ENTRE LA VARIABLE EQUIS Y -
C*                                LA CORRESPONDIENTE VARIABLE YE QUE SE
C*                                EVALUA A AQUÍ PERO OTRAS VECES SOLO SE
C*                                UTILIZAN LOS CALCULOS PREVIA-
C*                                MENTE DESARROLLADOS
C*
C*    VARIABLE IPAR               ENTRADA CONTENDRA LA INFORMACION
C*    ENTERA                      CORRESPON    DIENTE A CUANDO SE CALCULA LA
C*                                COVARIABILIDAD Y CUANDO NO POR EFECTO DE
C*                                AHORRO DE  PROCESADOR
C*
C*   FORMA DE USO  -  CALL VAVER(RMU,VAR,F,NMUE,Y,X,N,VAR1)
C*
C*   SUBRUTINAS QUE EMPLEA   -  SXYPOB
C*
C*-------------------------------------------------------------------------------------
C*
      SUBROUTINE VAVER(RMU,VAR,F,NMUE,Y,X,N,VAR1,SVAR1,SXY,IPAR)

C*                    ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE

      INTEGER N,NMUE,IPAR

      DIMENSION RMU(2),VAR(2),Y(500),X(N),VAR1(3),SVAR1(3)

      REAL RMU,VAR,Y,X,VAR1,F,SVAR1,SXY

C*                    ESPECIFICACION DE ELEMENTOS LOCALES

      REAL C1,C2
C*
C*                    PRIMERA INSTRUCCION EJECUTABLE
```

```
      C1 = ( 1 - F ) / FLOAT( NMUE )
      C2 = VAR(1) / ( RMU(1) * RMU(1) )

      IF (IPAR .NE. 1 ) GO TO 1
      SXY = 0.0
      CALL SXYPOB(X,Y,N,RMU(1),RMU(2),SXY)
1     VAR1(1) = C1 * VAR(2)
      SVAR1(1) = SQRT( VAR1(1) )

      VAR1(2) = C1 * ( VAR(2) + C2 * RMU(2) *
     *     RMU(2) - 2 * RMU(2) * SXY / RMU(1)  )
      SVAR1(2) = SQRT( VAR1(2) )
      VAR1(3) = C1 * ( VAR(2) - ( SXY * SXY ) / VAR(1)
     *     )
      SVAR1(3) = SQRT( VAR1(3) )
2     CONTINUE
      RETURN
      END
```

APÉNDICE C.2

PROGRAMA PPT

Programa que simuló el muestreo aleatorio con probabilidad proporcional al "tamaño" de las unidades, contemplando los estimadores para la media, en los casos de sin y con reemplazo respectivamente, desarrollado en FORTRAN 77, se tomó como base el programa desarrollado por Flores, así que sólo se muestra el programa principal y las subrutinas "models", "resul1" y "mappt" que fueron los que tuvieron modificaciones, las subrutinas "gepobg", "tranfs", "ajusta" fueron canceladas para la realización de la simulación del presente trabajo.

```
C*                 PROGRAMA
C*
C*         DESARROLLADO POR JOSE ANTONIO FLORES  DIAZ
C*
C*        FECHA DE LA ULTIMA MODIFICACION  REALIZADA
C*
C*           POR NANCY BEATRIZ ZUNIAGA HIDALGO
C*
C*                8 DE SEPTIEMBRE DE 1996
C*
C*------------------------------------------------------------------------------------------------
C*
C*          P R O G R A M A   P R I N C I P A L
C*
C*------------------------------------------------------------------------------------------------
C*
C*ESTE PROGRAMA DESARROLLA LA SIMULACION DEL MUESTREO
C*
C*ALEATORIO DE   UNIDADES CON PROBABILIDADES PROPORCIONALES AL
C*
C*TAMANIO DE LAS UNIDADES Y CALCULA EL ESTIMADOR SECUENCIAL QUE
C*
C*PROPONE DES RAJ ASI COMO TAMBIEN EVALUA EL ESTIMADOR QUE
C*
C*CORRESPONDERIA A UN PPT CON REEMPLAZO. LOS ANTERIORES
C*
C*CALCULOS LOS HACE PARA LA MEDIA
C*
C*  LO YA REFERIDO SE APLICA A LOS MODELOS SIGUIENTES
C*
C*        1.- COV(X,Y)= 0.3
C*
C*        2.- COV(X,Y)= 0.5
```

```
C*
C*        3.- COV(X,Y)= 0.8
C*
C*
C*  LOS ARREGLOS QUE UTILIZA SON:
C*
C*     ARREGLO XPOB        ESTE ARREGLO CONTENDRA LOS VALORES
C*     REAL                DE LAS 'X'
C*
C*     ARREGLO NELEM       ESTE ARREGLO CONTENDRA LOS VALORES
C*     ENTERO              DE LOS TAMANIOS DE POBLACION QUE SE
C*                         ESTAN CONSIDERANDO
C*
C*     ARREGLO YPOB        ESTE ARREGLO CONTENDRA LOS VALORES
C*     REAL                DE LOS DIFERENTES MODELOS CONSIDERADO
C*
C*     ARREGLO RMUPOB      ESTE ARREGLO CONTENDRA LOS VALORES
C*     REAL                CORRESPONDIENTES A LAS MEDIAS PO-
C*                         BLACIONALES DE LOS DIFERENTES MODELOS
C*                         CONSIDERADOS
C*     ARREGLO  MCOV       ESTE ARREGLO CONTENDRA LOS VALORES DE LAS
C*     REAL                COVARIANZAS CORRESPONDIENTES A CADA MODELO
C*
C*     ARREGLO VARPOB      ESTE ARREGLO CONTENDRA LOS VALORES
C*     REAL                CORRESPONDIENTES A LAS VARIANZAS
C*                         POBLACIONALES DE LOS DIFERENTES MO-
C*                         DE LOS CONSIDERADOS
C*
C* FORMA DE UTILIZARSE  - NINGUNA PORQUE SE TRATA DEL PROGRAMA PRINCI
C*                         PAL
C*
C* SUBRUTINAS QUE EMPLEA
C*              - MODELS
C*
C*              - TOTPOB
```

```
C*
C*                  - RESUL1
C*
C*                  - MAPPT
C*
C*------------------------------------------------------------------------------------------------
C*
      IMPLICIT NONE

      DIMENSION XPO(500),NELEM(9),YPOB(500,3),RMUPOB(4),VARPOB(4)

     *,XPOB(500),TABLAS(6,3),MCOV(3)

      REAL XPO,XPOB,YPOB,RMUPOB,TABLAS,SUMX,SUMX2,VARPOB,MCOV

      INTEGER NELEM,I,J,DSEED,DSEED1,ICASO

C*                   PRIMERA INSTRUCCION EJECUTABLE

      DATA NELEM/30,50,70,100,200,350/

      DATA MCOV/0.3,0.5,0.8/

      OPEN(6,FILE="FINAL.")

      OPEN(7,FILE="MOMPPT1.")

      OPEN(8,FILE="MOMPPT2.")

      OPEN(9,FILE="MOMPPT3.")

      OPEN(3,FILE="MEDVARS.")

       DSEED = 12357987
```

```
      DSEED1 = DSEED
      DO 2 I= 1,6
        DO 3 J=1,3
          CALL MODELS(XPOB,YPOB,NELEM(I),MCOV(J),SUMX,SUMX2)
          CALL TOTPOB(YPOB,NELEM(I),RMUPOB,VARPOB,SUMX,SUMX2)
          CALL RESUL1(RMUPOB,VARPOB,NELEM(I),MCOV(J))
          CALL MAPPT(XPOB,YPOB,NELEM(I),RMUPOB,VARPOB,SUMX,
     *     DSEED1,J)
 3      CONTINUE
 2    CONTINUE
      CLOSE(6)
      CLOSE(7)
      CLOSE(8)
      CLOSE(9)
      CLOSE(3)
      CALL EXIT
      END
C*
C* SUBRUTINA MODELS
```

```
C*-----------------------------------------------------------------------------------------------------------
C*
C* ESTA SUBRUTINA CALCULA O EVALUA LOS DIFERENTES MODELOS QUE
C*
C* SE ESPECIFICARON YA EN EL PROGRAMA PRINCIPAL, ADEMAS EVALUA
C*
C* LA SUMA DE LAS X COMO TAMBIEN LA SUMA DE LAS X AL CUADRADO
C*
C*   LOS PARAMETROS DE ESTA SUBRUTINA SON LOS SIGUIENTES:
C*
C*    ARREGLO X            ENTRADA CONTENDRA LOS VALORES
C*    REAL                 DESESTANDARIZADOS
C*
C*    ARREGLO Y            SALIDA CONTENDRA LOS VALORES DE LOS
C*    REAL                 DIFERENTES MODELOS
C*
C*    ARREGLO MCOV         ENTRADA CONTENDRA EL VALOR DE LA
C*    REAL                 COVARIANZA DE      LOS DIFERENTES MODELOS
C*
C*    VARIABLE NR          ENTRADA CONTENDRA EL TAMANIO DE LA POBLA
C*    ENTERA               NUMERO DE ELEMENTOS DE LA POBLACION
C*
C*    VARIABLE SUMX        SALIDA CONTENDRA EL VALOR CORRESPONDIENTE A
C*    REAL                 LA SUMA DE LAS X
C*
C*    VARIABLE SUMX2       SALIDA CONTENDRA EL VALOR CORRESPONDIENTE A
C*    REAL                 LA SUMA DE LAS X AL CUADRADO
C*
C*   FORMA DE USO  - CALL MODELS(X,Y,NR,COV,SUMX,SUMX2)
C*
C*   SUBRUTINAS QUE EMPLEA -  CHFAC (IMSL)
C*
C*                     RNMVN (IMSL)
C*
C*                     RNSET (IMSL)
```

```
C*
C*                    UMACH (IMSL)
C*
C*------------------------------------------------------------------------------------------------------------------------------
C*
      SUBROUTINE MODELS(X,Y,NR,MCOV,SUMX,SUMX2)

C*                    ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE

      INTEGER NR,IRANK,K,LDR,LDRSIG,NOUT,LI,I,ISEED

      REAL COV(2,2),R(500,2),RSIG(2,2)

      REAL SUMX, SUMX2, X2, X, Y,MCOV

      EXTERNAL CHFAC,RNMVN,RNSET,UMACH

      DIMENSION  X(NR),Y(500)

      CALL UMACH(2, NOUT)

      LDR =500

      K= 2

      LDRSIG =2

      ISEED = 12357987

      COV(1,1)=1.0

      COV(1,2)=MCOV
```

```
COV(2,1)=COV(1,2)

COV(2,2)=COV(1,1)

C* OBTIENE LA FACTORIZACION DE CHOLESKY

CALL CHFAC(K,COV,2,0.00001,IRANK,RSIG,LDRSIG)

CALL RNSET(ISEED)

CALL RNMVN(NR.K,RSIG,LDRSIG,R,LDR)

SUMX=0.0

SUMX2=0.0

DO 12 I=1,NR

C* SE INCREMENTA UNA CONSTANTE C=10 A LOS DATOS PARA EVITAR VALORES
C* NEGATIVOS EN LAS VARIABLES `X'

X(I)=R(I,1)+10

C*              CALCULA X AL CUADRADO

X2= X(I) * X(I)

C*              CALCULA LA SUMA DE LAS X

SUMX=SUMX+ X(I)

C*              CALCULA LA SUMA DE LAS X AL CUADRADO

SUMX2=SUMX2+ X2
```

```
      Y(I)=R(I,2)

12  CONTINUE

      RETURN

      END

C* SUBRUTINA RESULI
C*
C*------------------------------------------------------------------------------------------------------------
C*
C* ESTA SUBRUTINA SE ENCARGA DE IMPRIMIR LOS RESULTADOS CORRESPON-
C*
C* DIENTES A LA MEDIA Y A LA VARIANZA POBLACIONALES DE LAS VARIA-
C*
C* BLES AUXILIAR Y DEL RESTO DE LOS MODELOS Y'ES CONSIDERADOS
C*
C* LOS PARAMETROS QUE EMPLEA SON :
C*
C*
C*    ARREGLO RMUPOB        ENTRADA CONTENDRA LA INFORMACION DE ME-
C*    REAL                  DIAS PARA LA POBLACION TANTO DE SU CARAC-
C*                    TERISTICA AUXILIAR COMO LA DE LOS MODELOS
C*                          QUE SE ESTAN CONSIDERANDO
C*
C*    ARREGLO VARPOB        ENTRADA CONTENDRA LA INFORMACION DE VA-
C*    REAL                  RIANZAS PARA LA POBLACION TANTO DE SU
C*                          CARACTERISTICA AUXILIAR COMO LA DE LOS MO-
C*                          DE LOS QUE SE ESTAN CONSIDERANDO
C*
```

```
C*   VARIABLE MCOV        ENTRADA CONTENDRA EL VALOR DE COVARIANZA
C*   REAL                 DE LOS DIFERENTES MODELOS
C*   VARIABLE N           ENTRADA CONTENDRA LA INFORMACION DE EL
C*   ENTERA               NUMERO DE ELEMENTOS QUE COMPONEN LA PO-
C*                        BLACION

C* LA FORMA DE USO   -CALL RESUL1(RMUPOB,VARPOB,N,COV)
C*
C* SUBRUTINAS QUE EMPLEA  - NINGUNA
C*
C*-----------------------------------------------------------------------
C*
      SUBROUTINE RESUL1(RMUPOB,VARPOB,N,MCOV)

C*                  ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE
C*
      DIMENSION RMUPOB(2),VARPOB(2),TIT(2)

      REAL RMUPOB,VARPOB,TIT,RMUS,MCOV

      INTEGER N

C*                  ESPECIFICACION DE ELEMENTOS LOCALES

      INTEGER J,K

      TIT(1)= "X"

      TIT(2)= "Y"

C*                  PRIMERA INSTRUCCION EJECUTABLE

      WRITE(6,200) N, MCOV
```

```
      WRITE(6,201)
      DO 1 K = 1 , 2
         WRITE(6,202)TIT(K),RMUPOB(K),VARPOB(K)
    1 CONTINUE
      WRITE(6,203)
      RETURN
  200 FORMAT(1H1,11X,"T A B L A   D E   V A L O R E S   P O B L A",
     *"C I O N A L E S",//,36X,"( N = ",I3," )",//,36X,"(COV=",F5.3,")",
     *//,10X,50("-"))
  201 FORMAT(10X,"I MODELO O CONCEPTO  I",4X,"MEDIA",4X,"I",2X,"VA",
     *"RIANZA",3X,"I")
  202 FORMAT(10X,"I ",9X,A1,8X," I ",F11.5," I ",F11.5," I")
  203 FORMAT(10X,50("-"))
      END

C* SUBRUTINA MAPPT
C*
C*-----------------------------------------------------------------------------------------------
C*
C* ESTA SUBRUTINA SE ENCARGA DEL PRINCIPAL PROCESO EN LA SIMULACION
```

```
C*
C* DADO QUE EN ELLA SE REALIZA TODO LO REFERENTE A SELECCIONAR UNA
C*
C* MUESTRA CON PROBABILIDAD PROPORCIONAL AL "TAMANIO" SIN REEMPLAZO
C*
C*OBTENER ESTIMACIONES PARA LA MEDIA (SE PUEDE TAMBIEN PARA EL TOTAL)
C*
C*DE  ACUERDO AL CALCULO QUE PROPONE DES RAJ (ESTIMADOR SECUENCIAL) Y
C*
C* CON BASE A LA MUESTRA ELEGIDA SE DESARROLLAN LOS CALCULOS CORRES-
C*
C* PONDIENTES A UNA MUESTRA PPT CON REEMPLAZO QUE CORRESPONDERIA PARA
C*
C* LAS MISMAS UNIDADES PREVIAMENTE ELEGIDAS
C*
C* FINALMENTE CUANDO SE HAYA LLEGADO AL NUMERO DE MUESTRAS QUE SE
C*
C*DESEAN OBTENER ENTONCES SE PROSIGUE CON OTRO PORCENTAJE DE
C*
C* MUESTREO O SE REGRESA AL PROGRAMA PRINCIPAL
C*
C* TODO LO ANTERIOR LO DESARROLLA PARA LOS PORCENTAJES DE
C*
C* MUESTREO SIGUIENTES 10, 20, 30, 40 Y 45
C*
C* LOS PARAMETROS QUE EMPLEA SON :
C*
C*    ARREGLO X          ENTRADA CONTENDRA LA INFORMACION DE LA
C*    REAL               POBLACION AUXILIAR
C*
C*    ARREGLO Y          ENTRADA CONTENDRA LA INFORMACION DE LOS
C*    REAL               DIFERENTES MODELOS PARA LA POBLACION EN
C*                       ESTUDIO
C*
C*    VARIABLE N         ENTRADA CONTENDRA LA INFORMACION DE EL
```

```
C*    ENTERA                 NUMERO DE ELEMENTOS QUE COMPONEN LA PO-
C*                           BLACION
C*
C*    ARREGLO RMUPOB         ENTRADA CONTENDRA LA INFORMACION DE ME-
C*    REAL                   DIAS PARA LA POBLACION TANTO DE SU CARAC-
C*                           TERISTICA AUXILIAR COMO LA DE LOS MODELOS
C*                           QUE SE ESTAN CONSIDERANDO
C*
C*    ARREGLO VARPOB         ENTRADA CONTENDRA LA INFORMACION DE VA-
C*    REAL                   RIANZAS PARA LA POBLACION TANTO DE SU
C*                           CARACTERISTICA AUXILIAR COMO LA DE LOS MO-
C*                     DE LOS QUE SE ESTAN CONSIDERANDO
C*
C*    VARIABLE DSEEMA        ENTRADA/SALIDA CONTENDRA LA INFORMACION
C*    ENTERA                 CORRESPONDIENTE PARA QUE LAS SUBRUTINAS
C*                           DEL IMSL SE USEN
C*
C*    VARIABLE IMPRI         ENTRADA CONTENDRA LA INFORMACION QUE
C*    ENTERA                 DETERMINA CUANDO ESCRIBIR EN EL ARCHI-
C*                           VO DOS (N/NMUE)
C*
C*    VARIABLE SUMX          ENTRADA CONTENDRA EL TOTAL QUE CORRESPON-
C*                           DE A LA SUMA DE LOS VALORES DE LA VARIABLE
C*                           AUXILIAR DEL TOTAL DE LA POBLACION
C*
C*       ++++++++++++++++++++++++++++++++++++++++
C*
C* OTROS ARREGLOS O ELEMENTOS QUE SE USAN EN LA SUBRUTINA SON
C*
C*    ARREGLO PORMUE         SENIALA LOS PORCENTAJES DE MUESTREO
C*    REAL
C*
C*    ARREGLO IAPUNT         CONTENDRA LA INFORMACION DE LOS
C*    ENTERO                 INDICES DE LA POBLACION QUE SON ELEGIDOS
C*                           EN UNA MUESTRA
```

```
C*
C*   ARREGLO ESTM          CONTENDRA LA MEDIA ESTIMADA BAJO CADA ES-
C*   REAL                  QUEMA PARA LOS DIFERENTES MODELOS QUE SE
C*                    ESTAN CONTEMPLANDO
C*
C*   ARREGLO ESTV          CONTENDRA LA VARIANZA ESTIMADA BAJO CADA
C*   REAL                  ESQUEMA PARA LOS DIFERENTES MODELOS QUE SE
C*                         ESTAN COMTEMPLANDO
C*
C*   ARREGLO ARSUMX        ENTRADA CONTENDRA LA INFORMACION CORRES-
C*   REAL                  PONDIENTE  A LA SUMA DE LA VARIABLE AUXI-
C*                         LIAR PARA LAS UNIDADES QUE NO HABIAN SIDO
C*                         ELEGIDAS HASTA QUE FUE SELECCIONADO EL ELE
C*                              MENTO I
C*
C*     ++++++++++++++++++++++++++++++++++++++++++++++++
C*
C* FORMA DE USO   - CALL MAPPT(X,Y,N,RMUPOB,VARPOB,SUMX,DSEEMA,J)
C*
C*
C* SUBRUTINAS QUE EMPLEA :
C*
C*             -  GGUBFS  (IMSL)
C*
C*             -  UNE
C*
C*             -  SACA
C*
C*             -  EDRSEC
C*
C*             -  EWRPPT
C*
C*             -  RECUP
C*
C*             -  MEZCLA
```

```
C*
C*              - RNGET (IMSL)
C*
C*-----------------------------------------------------------------------------------------------
C*
      SUBROUTINE MAPPT(X,Y,N,RMUPOB,VARPOB,SUMX,DSEEMA,J)

C*                ESPECIFICACION DE ARGUMENTOS

      IMPLICIT NONE

      INTEGER N,DSEEMA

      DIMENSION X(N),Y(500),RMUPOB(2),VARPOB(2)

      REAL X,Y,RMUPOB,SUMX,VARPOB

C*                ESPECIFICACION DE ELEMENTOS LOCALES

      DIMENSION PORMUE(5),IAPUNT(500),ARSUMX(500),
     *ESTM(2),ESTV(2)

      REAL PORMUE,ARSUMX,ESTM,ESTV,TNMUE,SUMXP,RNUNF,FX,XM

      INTEGER IAPUNT,IENS,NMUE,UAPUXM,IAPUX,I,K,IAPUXM,J,M

      EXTERNAL RNUNF

C*                PRIMERA INSTRUCCION EJECUTABLE

      DATA PORMUE/.1,.2,.3,.4,.45/

      CALL RNSET(DSEEMA)
```

```
CALL UNE(IAPUNT,N,IAPUX)

DO 3 I = 1 , 5

   IENS = 0

   TNMUE = PORMUE(I) * FLOAT( N )

   NMUE = AINT( TNMUE )

   IF( TNMUE - NMUE .GT. 0.00000001 ) NMUE = NMUE + 1

   SUMXP = SUMX

   K = 0

1   FX = RNUNF()

   CALL RNGET(DSEEMA)

   XM = ( FLOAT( DSEEMA ) / ( 2147483647. / ( SUMXP + 1.0 )))

   IF(XM .GT. SUMXP) GO TO 1

   CALL SACA(X,IAPUNT,ARSUMX,N,IAPUX,IAPUXM,UAPUXM,XM,SUMXP)

   K = K + 1

   IF(K .LT. NMUE) GO TO 1

   IENS = IENS + 1


      CALL EDRSEC(X,Y,IAPUNT,ARSUMX,N,NMUE,IAPUXM,
```

```
*     ESTM(1),ESTV(1))
      CALL EWRPPT(X,Y,IAPUNT,ARSUMX,N,NMUE,IAPUXM,
*     ESTM(2),ESTV(2))
         K = J + 6
      WRITE(K,202) I,IENS,(ESTM(M),M=1,2),(ESTV(M),
*        M=1,2)

    CALL RECUP(IAPUNT,N,IAPUXM)
    CALL MEZCLA(IAPUNT,N,IAPUX,IAPUXM)
    K = 0

    SUMXP = SUMX
    IF(IENS .LT. 200) GO TO 1
 3  CONTINUE
   CALL RNGET(DSEEMA)
   RETURN
202 FORMAT(I1,I3,4F11.6)
   END
```

APÉNDICE C.3

PROGRAMAS DE SPSS

El primero sirvió para obtener estadísticas descriptivas de los estimadores.

El segundo para obtener estadísticas descriptivas de las estimaciones de las varianzas de los estimadores.

El tercero, sirvió para estandarizar los valores de los estimadores para todas y cada una de las 200 muestras desarrolladas en cada caso.

El cuarto, con los valores estandarizados de cada estimador, obtenidos con el programa anterior, se aplicó la prueba de Kolmogorov-Smirnov.

PROGRAMA QUE CALCULO FRECUENCIAS PARA LA MEDIAS

```
SET LENGTH 120.

SET WIDTH 80.

TITLE 'MUESTREO PPT MEDIAS'.

DATA LIST FILE = 'momppt1.nan' FIXED RECORDS = 1 NOTABLE

  /VAR1 1 VAR2 2-4 VAR3 5-17 VAR4 18-30 VAR5 31-43 VAR6 44-56
   VAR7 57-69 VAR8 70-83 .

VARIABLE LABELS VAR1 'FRACCION DE MUESTREO' /VAR2 'NUMERO DE
MUESTRA'
  /VAR3 'MEDESTDES-RAJ' /VAR4 'MEDESTPPTWR'
  /VAR5 'VARESTDES-RAJ' /VAR6 'VARESTPPTWR'
  /VAR7 'COEESTDES-RAJ' /VAR8 'COEESTPPTWR'.

SPLIT FILE BY VAR1.

FREQUENCIES VARIABLES = VAR3 TO VAR4 VAR7 VAR8 /FORMAT =
NOTABLE
  /STATISTICS = MEAN VARIANCE MINIMUM MAXIMUM MODE MEDIAN.

FINISH.
```

PROGRAMA QUE CALCULA LA FRECUENCIA DE LAS VARIANZAS

```
SET LENGTH 120.

SET WIDTH 80.

TITLE 'DATOS DE MUESTREO PPT VARIANZAS'.

DATA LIST FILE = 'momppt3.nan' FIXED RECORDS = 1 NOTABLE

 /VAR1 1 VAR2 2-4 VAR3 5-17 VAR4 18-30 VAR5 31-43
  VAR6 44-56  VAR7 57-69 VAR8 70-83.

VARIABLE LABELS VAR1 'NIVEL DE MUESTREO'

 /VAR2 'NO. DE MUESTRA'
 /VAR3 'MEDESTDES-REM' /VAR4 'MEDEST-DESRAJ'
 /VAR5 'VARESTDES-REM' /VAR6 'VAREST-DESRAJ'
 /VAR7 'COEESTDES-REM' /VAR8 'COEEST-DESRAJ'.

SPLIT FILE BY VAR1.

FREQUENCIES VARIABLES = VAR5 TO VAR6/FORMAT = NOTABLE
 /STATISTICS = MEAN VARIANCE MINIMUM MAXIMUM MODE MEDIAN.

FINISH.
```

PROGRAMA QUE ESTANDARIZO MEDIAS

```
SET LENGTH  120.

SET WIDTH 80.

TITLE 'MUESTREO PPT ESTADISTICAS PARA MEDIAS (CON REEMP. Y DES
RAJ)'.

DATA LIST FILE ="momppt1.nan" FIXED RECORDS = 1 NOTABLE
 /VAR1 1 VAR2 2-4 VAR3 5-17 VAR4 18-30 VAR5 31-43
  VAR6 44-56 VAR7 57-69 VAR8 70-83.

VARIABLES LABELS VAR1 'NIVEL DE MUESTREO'/ VAR2 'NO. DE MUESTRA'
 / VAR3 'MEDESTPPTWR' / VAR4 'MEDESTDES-RAJ' / VAR5 'VARESTPPTWR'
 / VAR6 'VARESTDES-RAJ' /  VAR7 'COEESTPPTWR' / VAR8 'COEESTDES-RAJ'.

SPLIT FILE BY VAR1.

DESCRIPTIVES VAR1 TO VAR8 /SAVE.

COMMENT OPTION 3.

SAVE OUTFILE = 'ACTVFILE' /KEEP = VAR1 VAR3 VAR4 ZVAR3
                    ZVAR4.

FINISH.
```

PROGRAMA QUE REALIZO LA PRUEBA DE KOLMOGOROV-SMIRNOV

```
SET LENGTH 120.

SET WIDTH 80.

TITLE 'MUESTREO NOPPT MEDIAS'.

FILE HANDLE ACTVFILE /NAME = 'ACTVFILE'.

GET FILE = ACTVFILE.

SPLIT FILE BY VAR1.

NPAR TEST K-S (NORMAL,0,1) = ZVAR3 TO ZVAR4.

FINISH.
```

BIBLIOGRAFÍA.

1. CANAVOS George C. "*Probabilidad y Estadística Aplicaciones y Métodos*" McGRAW-HILL, México; 1991.

2. COCHRAN William G. "*Técnicas de Muestreo*" De. CECSA; Novena reimpresión; México, 1992.

3. CONOVER W. J. "*Practical Nonparametric Statistics*" John Wiley and Sons; New York, USA; 1980.

4. COURANT R; Fritz John. "*Introducción al calculo y al análisis matemático*" Limusa; México D.F; 1998.

5. DOMETRIUS Nelson C. "*Social Statistics Using SPPSS*" HarperCollins Publisher Inc; 1992.

6. FLORES DIAZ José Antonio, "*Comparación de esquemas muestrales con y sin reemplazo*", Tesis para obtener el título de M. en C; 1992.

7. IMSL. "*User's Manual Stat/Library*" IMSL, Inc; Houston, Texas, USA; 1987.

8. KISH Leslie. "*Muestreo de Encuestas*" Trillas; México; 1972.

9. MANUAL. "*Guía introductoria a la computadora CRAY Y-MP*" Computo académico UNAM.

10.McCRACKEN, Danie. "*FORTRAN 77*' Limusa, S.A. México D.F; decimocuarta impresión 1976.

11.MENDENHALL William; Scheaffer Richard L; Wackely Dennis. "*Estadística Matemática con Aplicaciones*" University of Florida; Iberoamérica, S. A. de C. V; México; 1991.

12.MOOD, Alexander; A. Graybill, Franklin; C. Boes, Duane. "*Introduction to the Theory of Statistics*" McGraw-Hill, Kogakusha, LTD; Tokio, Japan;Third Edition; 1974.

13.ORGANICK, Elliot I. "*FORTRAN IV*" Fondo educativo interamericano, S.A;Colombia.

14.RAJ, Des. "*Sample Theory*" McGraw-Hill; New York, USA; 1968.

15.SAHAI Hardes; Martínez Wilfredo. "*Tablas y Fórmulas Estadísticas para las ciencias Biológicas, Sociales y Físicas*" Iberoaméricana S. A. de C. V; 1996.

16.SCHEAFFER Richard L; Mendenhall William; Ott Lyman. "*Elementos de Muestreo*" Iberoaméricana, S.A: de C. V; México; 1993.

17.SEBER, G. A. F. "*Multivariate Observation*" John Wiley and Son; New York; 1984.

18.SPSSX. "*User's Guide*" SPSS, Inc; Chicago, Illinois, USA.; 1989.

19.USER'S MANUAL. "*International Mathematical and Statistical Librarien*" Fortran Subroutine for Mathematics and Statistics, Edición 9°, Noviembre 1984.