5195
REH

CIB - ESPOL

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

INSTITUTO DE CIENCIAS MATEMÁTICAS
INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

" ESTUDIO COMPARATIVO, POR SIMULACION DE TECNICAS DE MUESTREO ALEATORIAS Y SISTEMATICAS"

TESIS DE GRADO

PREVIO A LA OBTENCION DEL TITULO DE:

INGENIERO EN ESTADISTICA INFORMATICA

PRESENTADA POR:

LUDGARDA REHFISCH PAREDES

GUAYAQUIL – ECUADOR

1999

CIB
D-I 9588

ESPOL
1958
ESCUELA SUPERIOR
POLITECNICA DEL LITORAL
CIB - ESPOL

ESPOL
1958
ESCUELA SUPERIOR
POLITECNICA DEL LITORAL
CIB - ESPOL

AGRADECIMIENTO

A Dios por la fuerza espiritual que me ha brindado cada día para seguir adelante.

Al Ing. Zurita y a todas aquellas personas que colaboraron en la realización de esta tesis.

ESPOL
1958
ESCUELA SUPERIOR
POLITECNICA DEL LITORAL
CIB - ESPOL

DEDICATORIA

A mis padres por todo el apoyo que me dieron a lo largo de mi vida.

A Verónica y a David, por toda la confianza que depositaron en mí.

A Marcelo, porque siempre estuvo presente en los momentos difíciles.

TRIBUNAL DE GRADUACION

.....

Ing. Félix Ramírez Cruz

Director del ICM

.....

Ing. Gaudencio Zurita Herrera

Director de Tesis

.....

Ing. Washington Armas Cabrera

Vocal

.....

Ing. Francisco Torres Andrade

Vocal

DECLARACION EXPRESA

"La responsabilidad por todos los hechos, ideas y doctrinas expuestos en este trabajo, me corresponde exclusivamente; y, el patrimonio intelectual de la misma, a la ESCUELA SUPERIOR POLITECNICA DEL LITORAL"

(Reglamento de Exámenes y Títulos profesionales de la ESPOL)

Rehfisch P.

.....

Srta. Ludgarda Rehfisch Paredes

RESUMEN

El Muestreo es un conjunto de técnicas estadísticas que nos permite obtener información sobre las características de un conjunto a partir de una fracción de este.

En la primera parte de este trabajo realizamos una introducción de lo que es el Muestreo Probabilístico y de las ventajas que tiene en comparación con los censos completos, así también introducimos algunas definiciones importantes en este contexto y describimos cada una de las técnicas de Muestreo que vamos a utilizar.

En la segunda parte, analizamos las propiedades que tienen los estimadores de la media poblacional y del total que se obtienen con diferentes tipos de Muestreo.

A continuación, realizamos una verificación por simulación de las propiedades analíticas de los estimadores de la media, del total y de la proporción, así como también analizamos la cobertura que tienen los intervalos de confianza construidos para estos estimadores con las diferentes técnicas. Se trabajará con poblaciones especificas.

Finalmente, efectuamos un análisis de algunos estimadores de la media y del total, pero utilizando poblaciones mezcladas, tanto discretas como continuas.

INDICE GENERAL

Página

Resumen VI

Indice General VII

Indice de Figuras X

Indice de Tablas XV

1. FUNDAMENTOS DEL MUESTREO ALEATORIOS

SIMPLE Y SISTEMATICO 18

1.1 Introducción.. 18

1.2 Definiciones y Resultados Preliminares 20

1.3 Muestreo Aleatorio Simple 35

1.3.1 Propiedades Deseables de un Estimador 37

1.3.2 Estimaciones 40

1.3.3 Estimación para proporciones 52

1.4 Muestreo Estratificado 58

1.4.1 Estimaciones 59

1.5 Muestreo Sistemático 67

2.COMPARACION DE LAS PROPIEDADES
ANALITCAS DEL MUESTREO ALEATORIO
SIMPLE, ESTRATIFICADO Y SISTEMATICO. 72

2.1 Generalidades 72

2.2. Comparación de los Resultados Obtenidos con las
diferentes técnicas de muestreo 74

3. VERIFICACION DE LAS PROPIEDADES
ANALITICAS DE LOLSO DIFERENTES
TIPOS DE MUESTREO UTILIZNDO
SIMULACIÓN ESTOCASTICA. 87

3.1 Análisis con los resultados obtenidos con el
Muestreo Aleatorios Simple y con el Muestreo
Sistemático 90

3.2 Análisis con los resultados obtenidos con el
Muestreo Aleatorios Simple, con el Muestreo
Sistemático y con el Muestreo Estratificado.. . . . . 10 1

ESPOL
1970
ESCUELA SUPERIOR
POLITECNICA DEL LITORAL
CIB - ESPOL

4. COMPARACION DE LOS RESULTADOS OBTENIDOS CON LOS DIFERENTES TIPOS DE MUESTREO UTILIZANDO POBLACIONES DISCRETAS Y CONTINUAS. 118

4.1 Generalidades 118

4.2 Análisis con los resultados obtenidos con el Muestreo Aleatorios Simple y con el Muestreo Sistemático, utilizando una población Discreta 121

4.3 Análisis con los resultados obtenidos con el Muestreo Aleatorios Simple y con el Muestreo Sistemático utilizando una población continua..... 135

CONCLUSIONES Y RECOMENDACIONES.. 158

Conclusiones 158

Recomendaciones 160

ANEXOS 161

ANEXO 1.. 162

ANEXO 2 168

BIBLIOGRAFÍA. 177

INDICE DE FIGURAS

Página

FIGURA 1.1 Histograma de Probabilidades para $X_{(1)}$ 29

FIGURA 1.2 Histograma de Probabilidades para $X_{(2)}$ 30

FIGURA13 Histograma de Probabilidades para $X_{(3)}$ 31

FIGURA 1.4 Histograma de Probabilidades para $\overline{X}$ 31

FIGURA 1.5 Histograma de Probabilidades para $\tilde{X}$ 32

FIGURA 1.6 Histograma de Probabilidades para los Ingresos de la PEA 49

FIGURA 1.7 Composición de las k muestras. 68

FIGURA 1.8 Muestras seleccionadas 70

FIGURA 2.1 Histograma de Frecuencia para la Media Aritmética.(Muestreo Aleatorio Simple, n=437) 77

FIGURA 2.2 Histograma de Frecuencia para la Media Estratificada (Muestreo Estratificado, n=43 7) 78

FIGURA 2.3 Histograma de Frecuencia para la Media Sistemática. (Muestreo Sistemático, n=43 7) 79

FIGURA 2.4 Varianzas de los Estimadores..... 80

FIGURA 2.5 Distribuciones de las Medias, n=437..... 81

FIGURA 2.6 Histograma de Frecuencia para el Total

(Muestreo Aleatorio Simple, n=437) 83

FIGURA 2.7 Histograma de Frecuencia para el Total

(Muestreo Estratificado, n=437)..... 83

FIGURA 2.8 Histograma de Frecuencia para la

Media Aritmética (Muestreo

Sistemático, n=43 7) 84

FIGURA 2.9 Distribución del Total, n=437..... 8.5

FIGURA 3.1 Histograma de Frecuencia para la

Media Aritmética (Muestreo

Aleatorio Simple, n=30) 94

FGURA 3.2 Histograma de Frecuencia para

la Media Sistemática (Muestreo

Sistemático, n=l OO) 95

FIGURA 3.3 Varianza e la Media Aritmética

N(2,10),n=30. 96

FIGURA 3.4 Varianza e la Media Sistemática

N(2,10),n=30 96

FIGURA 3.5 Histograma de Frecuencia para el Total (Muestreo Aleatorio Simple, n=30) 97

FIGURA 3.6 Histograma de Frecuencia para el Total (Muestreo Sistemático, n=30) 98

FIGURA 3.7 Histograma de Frecuencia para la Proporción (Muestreo Aleatorio Simple, n=30) 99

FIGURA3.8 Histograma de Frecuencia para la Proporción (Muestreo Sistemático, n=30) 99

FIGURA39 Histograma de Frecuencia para la Media Aritmética. (Muestreo Aleatorio Simple, n=100) 106

FIGURA 3.10 Histograma de Frecuencia para la Media Sistemática. (Muestreo Sistemático,n= 1 OO) 107

FIGURA 3.11 Histograma de Frecuencia para la Media Estratificada. (Muestreo Estratificado, n= 1 OO) 107

FIGURA 3.12 Varianza de la Media Aritmética, n=l OO 108

FIGURA3.13 Varianza de la Media Estratificada, n=l OO..... 109

FIGURA3.14 Varianza de la Media Sistemática,.n=100 110

FIGURA 3.15 Histograma de Frecuencia para el Total, n= 1 OO (Muestreo Aleatorio Simple) ..... ll 1

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

FIGURA 3.16 Histograma de Frecuencia para el Total, n= 1 OO (Muestreo Estratificado) 112

FIGURA 3.17 Histograma de Frecuencia para el Total, n=100 (Muestreo Sistemático) 1 13

FIGURA 3.18 Histograma de Frecuencia para la Proporción, (Muestreo Aleatorio Simple, n=100) 1 14

FIGURA 3.19 Histograma de Frecuencia para la Proporción, (Muestreo Estratificado, n=1 OO) 114

FIGURA 3.20 Histograma de Frecuencia para la Proporción, (Muestreo Sistemático, n=1 OO) 115

FIGURA 4.1 Histograma de Frecuencias para la Media Aritmética, para la Media Estratificada y para la Media Sistemática, n=100 125

FIGURA 4.2 Varianza de los Estimadores 126

FIGURA 4.3 Histograma de Frecuencias para x, x_{sist} , $\bar{x}_{est}$ 138

FIGURA 4.4 Varianza de los Estimadores de la Media 139

FIGURA 4.5 Histograma de Frecuencia para el Total (Muestreo Aleatorio Simple,n=100) 145

FIGURA 4.6 Histograma de Frecuencia para el Total (Muestreo Sistemático,n=1OO) 145

FIGURA 4.7 Histograma de Frecuencia para el Total

(Muestreo Estratificado, n= 1 OO) 146

FIGURA 4.8 Histograma de Frecuencia para x, x_{sist} , $\bar{x}_{est}$ 149

FIGURA 4.9 Varianza del Estimador de la Media 150

FIGURA 4.10 Histograma de Frecuencia para el Total

(Muestreo Aleatorio Simple, n=100) 156

FIGURA 4.11 Histograma de Frecuencia para el Total

(Muestreo Sistemático, n=100) 156

FIGURA 4.12 Histograma de Frecuencia para el Total

(Muestreo Estratificado, n=100) 157

INDICE DE TABLAS

Página

TABLA 1 Muestras Posibles, N=9, n=3 25

TABLAII Distribuciones de X(l), X(2), X(3),X y $\tilde{X}$ 27

TABLAIII Medias, Varianzas, y F.G.M, N=9, n=3 34

TABLA IV Datos de la Ilustración 48

TABLAV Muestreo Aleatorio Simple: Proporciones 56

TABLAVI Cuadro Comparativo del estimador de la media poblacional, n=437. 75

TABLAVII Cuadro Comparativo del estimador del total poblacional, n=437. 76

TABLAVIII Comparación de los Resultados obtenidos con el Muestreo Aleatorio Simple y con el Muestreo Sistemático, n=30 92

TABLAIX Prueba K-S para las Medias $\bar{x}$ y para $\bar{x}_{sist}$ 93

TABLA X Comparación del Muestreo Aleatorio Simple, Estratificado, y Sistemático. (Mezcla de Poblaciones: N(2,10); N80,30) N(5,8), n=100) 103

TABLA XI Prueba K-S para las Medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$
(Mezcla de Poblaciones: N(2,10); N80,30),
N(5,8), n=100) 105

TABLA XII Cuadro Comparativo del estimador de la media
poblacional.
(Mezcla de Poblaciones: P(2), P(5), P(10), n=30)..... 123

TABLA XIII Prueba K-S para las Medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$
(Mezcla de Poblaciones: P(2), P(5), P(10), n=30). 127

TABLA X IV Cuadro Comparativo del estimador del total
poblacional.
(Mezcla de Poblaciones: P(2), P(5), P(10), n=30). 129

TABLA XV Comparación de los resultados obtenidos
con muestreo Aleatorio Simple y con el
sistemático., n=1 OO. 132

TABLA XVI Prueba K-S para las Medias $\bar{x}$ y $\bar{x}_{sist}$
(Poblaciones Discretas) 134

TABLA XVII Comparación de los resultados obtenidos
con Muestreo Aleatorio Simple, Muestreo
Estratificado y Muestreo Sistemático.
(Mezcla de Poblaciones: N(2,7), N(2,24),
N(2,13),n=100)..... 137

TABLA XVIII Prueba K-S para las Medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$ (Mezcla de Poblaciones: N(2,7), N(2,24), N(2,13),n=100) 142

TABLA XIX Comparación de los resultados obtenidos con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático,n=l OO (población Continua) 143

TABLAX X Comparación de los resultados obtenidos con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático,n= 1 OO (Mezcla de Poblaciones: N(4,15), N(2,15), N(8,15),n=100)..... 148

TABLA XXI Prueba K-S para las Medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$ (Mezcla de Poblaciones: N(2,7), N(2,24), N(2,13),n=100)..... 153

TABLAXXII Comparación de los resultados obtenidos con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático,n=100 (Mezcla de Poblaciones: N(2,7), N(2,24), N(2,13),n=100)..... 154

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

CAPITULO I

FUNDAMENTOS DEL MUESTREO ALEATORIO SIMPLE Y SISTEMATICO.

1.1. Introducción.

Los conocimientos, las actitudes y las acciones que tomamos se basan en gran parte sobre información contenida en fragmentos de un determinado fenómeno o hecho. Así por ejemplo, la opinión de una persona que labora en un banco, muchas veces se fundamenta en uno o más contactos que ha tenido durante su vida, de la misma manera, una persona estudiada en ciencias políticas de un país y que escribe un artículo periodístico se basa en las experiencias o en los conocimientos que él tiene cimentado en una información parcial; de ahí que tanto en la ciencia como en los aspectos humanos, en algunas ocasiones carecemos de los recursos necesarios para estudiar exhaustivamente los fenómenos que necesitamos conocer y que pueden ampliar o consolidar nuestro conocimiento sobre el mismo.

Por estas consideraciones, hace algún tiempo se desarrolló la teoría del Muestreo, la misma que tiene como propósito desarrollar métodos para la selección de

subconjuntos que proporcionen estimaciones con la suficiente “precisión” para los fines propuestos. Sobre la base de esta teoría, podemos obtener información sobre las características de un conjunto a partir de un fragmento de éste, es decir, ya no es necesario realizar una enumeración completa o censo del conjunto, salvo para ocasiones en las cuales se requiere información especifica de cada miembro del conjunto.

El Muestreo es un conjunto de técnicas estadísticas que exige muchas habilidades, en comparación con los censos y estudios exhaustivos, éste requiere menos trabajo material, pero más refinamiento y preparación.

Entre las ventajas que ofrece el muestreo se pueden citar: reducción del costo, mayor rapidez en la recolección de información y mayor precisión en los resultados obtenidos. Si la información se obtiene de una pequeña fracción del total, el costo asociado será menor que si se realizara un censo completo, por ejemplo en los Estados Unidos, algunas investigaciones importantes se las lleva a cabo utilizando una persona por cada 1300 elementos de la población, además cuando se toma solo una parte del total, se puede tener una mejor formación del personal y mayor supervisión del trabajo de campo y del procesamiento de los resultados, produciendo de esta manera resultados más precisos.

El Muestreo ha desempeñado un papel importante dentro de los censos nacionales, por ejemplo, en Estados Unidos en el censo de 1940 se tomó el 5% de la población para obtener información sobre ocupación, fertilidad, tamaño de la familia, etc. Desde 1950 el uso del muestreo aumento gradualmente y en nuestro país se intensifica su uso fundamentalmente en las encuestas de opinión, encuestas de hogares, encuestas industriales y agrícolas; estas últimas realizadas por el instituto Nacional de Estadísticas y Censos(INEC).

1.2. Definiciones y Resultados Preliminares.

Existe más de una forma de muestrear, pero antes de pasar a describir los métodos más habituales es necesario introducir algunas definiciones importantes en este contexto, tales como: universo, población, unidad de muestreo y muestra.

Universo

Es un conjunto bien definido de entes con una o más características sobre los cuales deseamos realizar algún tipo de inferencia.

Población.

Es un conjunto de valores cualitativos o cuantitativos que pueden tomar una característica del universo y los representamos por $x_1, x_2, x_3, \ldots x_N$, donde N es el tamaño de la población , o número de items en el universo. La característica se la denota usualmente por X y en teoría probabilística es denominada variable aleatoria.

Unidad de Muestreo

Es el objeto sobre el cual se efectúan las mediciones.

Muestra.

Es un subconjunto de tamaño n tomado de la población. Los elementos de la muestra son representados por $x_1, x_2, x_3, \ldots\ldots x_n$ $(n < N)$.

A continuación se presenta un ejemplo a través del cual se ilustran los conceptos antes enunciados.

Supongamos que se toma una circunscripción geográfica del país que tiene 100.000 elementos en su PEA (Población Económicamente Activa), y que se desea realizar un estudio acerca de los ingresos mensuales de los mismos.

El ***universo*** estará conformado por todos los elementos de la PEA en la provincia, mientras que la ***población*** estará constituida por los ingresos cada uno de ellos perciben.

Si para realizar este estudio se toma un 10% de la población, a esta fracción se la denominará muestra.

Para fines estadísticos una muestra, debe ser una parte "representativa" de la población. Para que una muestra sea lo más representativa posible, se utiliza la aleatoriedad en la selección de la misma, de esta manera se pretende lograr que el subconjunto escogido refleje las características de la población.

A continuación, dadas estas definiciones, se ilustra a través de un ejemplo la manera en la cual se puede obtener muestras a partir de una población finita, además, determinaremos las distribuciones que tienen algunas variables aleatorias derivadas de muestras que son importantes en inferencia estadística, como la media aritmética, la mediana muestra1 y los estadígrafos de orden, así también realizaremos otros cálculos como la media, la varianza y la función generadora de momentos(F.G.M), pero antes de realizar estos cálculos es necesario introducir la definición de Función Generatriz de Momentos.

Función Generatriz de Momentos

La función generatriz de momentos de una población (variable aleatoria)X, donde existe,

$$M_X(t) = E(e^{tx}) = \sum e^{tx} \cdot f(x) \quad ; \; t = (-h, h) \quad \text{cuando X es discreta y}$$

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \& ; \quad t = (-h, h) \quad \text{cuando X es continua.}$$

Siendo f(x) la distribución de Probabilidades de X.

Ilustración:

Supongamos que se tiene la siguiente población hipotética de tamaño N=9, dada por:

$$C = \{1,2,3,4,5,6,7,8,9\}$$

Donde cada elemento de la población X tiene igual probabilidad de ser elegido, esto es,

$$P(X=x) = \begin{cases} 1/9 & x = 1, 2, 3, 4, 5, 6, 7, 8, 9 \\ 0 & \text{el resto de x} \end{cases}$$

con media poblacional μ dada por

$$E[x] = \mu = \sum_{i=1}^{9} \frac{X_i}{9} = \frac{1+2+3+4+5+6+7+8+9}{9}$$

$$\mu = 5$$

varianza dada por,

$$\sigma^2 = E[(X-\mu)^2]$$

$$\sigma^2 = \sum_{x=1}^{9}(X-\mu)^2 P(X=x)$$

siendo $P(X=x) = \frac{1}{9}$

$$\sigma^2 = \frac{1}{9}((1-5)^2 + (2-5)^2 + (3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 +$$

$$(7-5)^2 + (8-5)^2 + (9-5)^2)$$

$$\sigma^2 = 6.66$$

y función generadora de momentos igual a

$$M_x[t] = \frac{1}{9}\left(e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t} + e^{7t} + e^{8t} + e^{9t}\right)$$

nótese que

$$\left.\frac{\partial M}{\partial t}\right/_{t=0} = \mu \qquad \sigma^2 = E[x^2] - \mu^2 = \left.\frac{\partial^2 M}{\partial t^2}\right/_{t=0} - \left(\left.\frac{\partial M}{\partial t}\right/_{t=0}\right)^2 = 6.66$$

Tomemos ahora, muestras de tamaño n = 3, como se sabe, el número de muestras posibles de tamaño n que se pueden tomar de una población de tamaño N, esta dado por,

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

en este caso tendremos 84 muestras posibles, (véase TABLA I)

TABLA I. Muestras Posibles, N=9,n=3

{1,2,3}	{1,3,7}	{1,5,8}	{2,3,6}	{2,5,7}	{3,4,6}	{3,6,9}	{4,6,9}	{6,7,8}
{1,2,4}	{1,3,8}	{1,5,9}	{2,3,7}	{2,5,8}	{3,4,7}	{3,7,8}	{4,7,8}	{6,7,9}
{1,2,5}	{1,3,9}	{1,6,7}	{2,3,8}	{2,5,9}	{3,4,8}	{3,7,9}	{4,7,9}	{6,8,9}
{1,2,6}	{1,4,5}	{1,6,8}	{2,3,9}	{2,6,7}	{3,4,9}	{3,8,9}	{4,8,9}	{7,8,9}
{1,2,7}	{1,4,6}	{1,6,9}	{2,4,5}	{2,6,8}	{3,5,6}	{4,5,6}	{5,6,7}	
{1,2,8}	{1,4,7}	{1,7,8}	{2,4,6}	{2,6,9}	{3,5,7}	{4,5,7}	{5,6,8}	
{1,2,9}	{1,4,8}	{1,7,9}	{2,4,7}	{2,7,8}	{3,5,8}	{4,5,8}	{5,6,9}	
{1,3,4}	{1,4,9}	{1,8,9}	{2,4,8}	{2,7,9}	{3,5,9}	{4,5,9}	{5,7,8}	
{1,3,5}	{1,5,6}	{2,3,4}	{2,4,9}	{2,8,9}	{3,6,7}	{4,6,7}	{5,7,9}	
{1,3,6}	{1,5,7}	{2,3,5}	{2,5,6}	{3,4,5}	{3,6,8}	{4,6,8}	{5,8,9}	

Una vez obtenidas todas las posibles muestras procedemos a calcular la media aritmética y la mediana de cada una de ellas para así determinar la distribución de probabilidades para el estadígrafo de orden 1 ($X_{(1)}$), el estadígrafo de orden 2 ($X_{(2)}$), el estadígrafo de orden 3 ($X_{(3)}$), para la media aritmética ($\overline{X}$) y para la mediana muestral($\tilde{X}$)

La mediana muestral estará dada por,

ESPOL
POLITECNICA DEL LITORAL
CIB - ESPOL

$$\tilde{X} = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{Si } n \text{ es impar} \\ \dfrac{X_{(n+2)} + X_{\left(\frac{n}{2}+1\right)}}{} & \text{Si n es par} \end{cases}$$

Mientras que $X_{(1)} \leq X_{(2)} \leq X_{(3)}$

A continuación se presentan los detalles de cada uno de los cálculos realizados.(ver TABLA II.) . En esta tabla no se muestra el cálculo de la mediana porque para nuestro ejemplo al ser el tamaño de las muestras igual a 3, la mediana coincide con el estadígrafo de orden 2.

Es así como para la muestra 1: $\{1,2,3\}$ se tiene que,

$$X_{(1)} = 1; \quad X_{(2)} = \tilde{X} = 2; \quad X_{(3)}=3; \quad \overline{X}=2$$

y para la muestra 84: $\{7,8,9\}$ se tiene que,

$$X_{(1)} = 7; \quad X_{(2)}=\tilde{X} = 8; \quad X_{(3)}=9; \quad \overline{X}=8$$

TABLA IL Distribución de $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, $\overline{X}$ y de $\tilde{X}$

Ordinal	Muestra	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$\overline{X}$
1	{1,2,3}	1	2	3	2
2	{1,2,4}	1	2	4	2,33
3	{1,2,5}	1	2	5	2,66
4	{1,2,6}	1	2	6	3
5	{1,2,7}	1	2	7	3,33
6	{1,2,8}	1	2	8	3,66
7	{1,2,9}	1	2	9	4
8	{1,3,4}	1	3	4	2,66
9	{1,3,5}	1	3	5	3
10	{1,3,6}	1	3	6	3,33
11	{1,3,7}	1	3	7	3,66
12	{1,3,8}	1	3	8	4
13	{1,3,9}	1	3	9	4,33
14	{1,4,5}	1	4	5	3,33
15	{1,4,6}	1	4	6	3,66
16	{1,4,7}	1	4	7	4
17	{1,4,8}	1	4	8	4,33
18	{1,4,9}	1	4	9	4,66
19	{1,5,6}	1	5	6	4
20	{1,5,7}	1	5	7	4,33
21	{1,5,8}	1	5	8	4,66
22	{1,5,9}	1	5	9	5
23	{1,6,7}	1	6	7	4,66
24	{1,6,8}	1	6	8	5
25	{1,6,9}	1	6	9	5,33

Ordinal	Muestra	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$\overline{X}$
26	{2,4,5}	2	4	5	3,66
27	{2,4,6}	2	4	6	4
28	{2,4,7}	2	4	7	4,33
29	{2,4,8}	2	4	8	4,66
30	{2,4,9}	2	4	9	5
31	{2,5,6}	2	5	6	4,33
32	{2,5,7}	2	5	7	4,66
33	{2,5,8}	2	5	8	5
34	{2,5,9}	2	5	9	5,33
35	{2,6,7}	2	6	7	5
36	{2,6,8}	2	6	8	5,33
37	{2,6,9}	2	6	9	5,66
38	{2,7,8}	2	7	8	5,66
39	{2,7,9}	2	7	9	6
40	{2,8,9}	2	8	9	6,33
41	{3,4,5}	3	4	5	4
42	{3,4,6}	3	4	6	4,33
43	{3,4,7}	3	4	7	4,66
44	{3,4,8}	3	4	8	5
45	{3,4,9}	3	4	9	5,33
46	{3,5,6}	3	5	6	4,66
47	{3,5,7}	3	5	7	5
48	{3,5,8}	3	5	8	5,33
49	{3,5,9}	3	5	9	5,66
50	{3,6,7}	3	6	7	5,33

Ordinal	Muestra	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$\overline{X}$
51	{1,7,8}	1	7	8	5,33
52	{1,7,9}	1	7	9	5,66
53	{1,8,9}	1	8	9	6
54	{2,3,4}	2	3	4	3
55	{2,3,5}	2	3	5	3,33
56	{2,3,6}	2	3	6	3,66
57	{2,3,7}	2	3	7	4
58	{2,3,8}	2	3	8	4,33
59	{2,3,9}	2	3	9	4,66
60	{4,6,7}	4	6	7	5,66
61	{4,6,8}	4	6	8	6
62	{4,6,9}	4	6	9	6,33
63	{4,7,8}	4	7	8	6,33
64	{4,7,9}	4	7	9	6,66
65	{4,8,9}	4	8	9	7
66	{5,6,7}	5	6	7	6
67	{5,6,8}	5	6	8	6,33

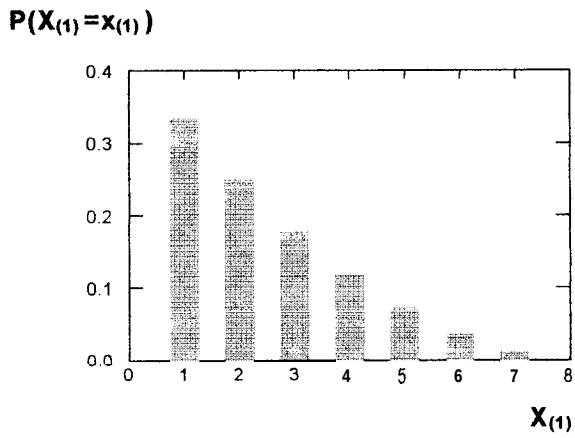
Ordinal	Muestra	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$\overline{X}$
68	{3,6,8}	3	6	8	5,66
69	{3,6,9}	3	6	9	6
70	{3,7,8}	3	7	8	6
71	{3,7,9}	3	7	9	6,33
72	{3,8,9}	3	8	9	6,66
73	{4,5,6}	4	5	6	5
74	{4,5,7}	4	5	7	5,33
75	{4,5,8}	4	5	8	5,66
76	{4,5,9}	4	5	9	6
77	{5,6,9}	5	6	9	6,66
78	{5,7,8}	5	7	8	6,66
79	{5,7,9}	5	7	9	7
80	{5,8,9}	5	8	9	7,33
81	{6,7,8}	6	7	8	7
82	{6,7,9}	6	7	9	7,33
83	{6,8,9}	6	8	9	7,66
84	{7,8,9}	7	8	9	8

Basados en los cálculos previos, se presenta las distribuciones de probabilidad para $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, $\overline{X}$ y para $\tilde{X}$.

Distribución de probabilidades de $X_{(1)}$:

$$P(X_{(1)} = x_{(1)}) = \begin{cases} 28/84 & X_{(1)}=1 \\ 21/84 & X_{(1)}=2 \\ 15/84 & X_{(1)}=3 \\ 10/84 & X_{(1)}=4 \\ 6/84 & X_{(1)}=5 \\ 3/84 & X_{(1)}=6 \\ 1/84 & X_{(1)}=7 \\ \\ 0 & \text{el resto de } X_{(1)} \end{cases}$$

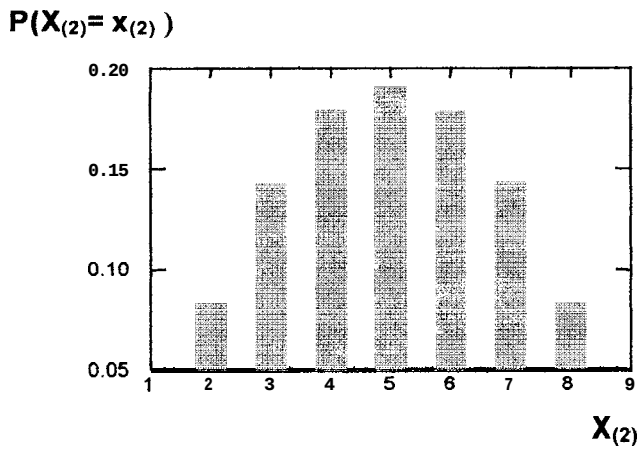
FIGURA 1.1. Histograma de Probabilidades de $X_{(1)}$



Distribución de probabilidades de $X_{(2)}$:

$$P(X_{(2)} = x_{(2)}) = \begin{cases} 7/84 & X_{(2)}=2,\ 8 \\ 12/84 & X_{(2)}=3,\ 7 \\ 15/84 & X_{(2)}=4,\ 6 \\ 16/84 & X_{(2)}=5 \\ 0 & \text{el resto de } X_{(2)} \end{cases}$$

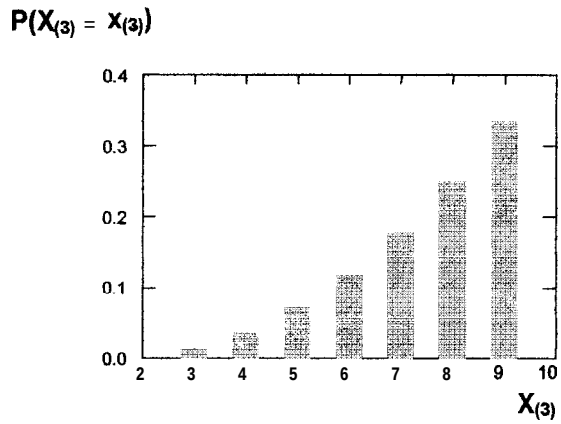
FIGURA 1.2. Histograma de Probabilidades de $X_{(2)}$



Distribución de probabilidades de $X_{(3)}$:

$$P(X_{(3)} = x_{(3)}) = \begin{cases} 1/84 & X_{(3)}=3 \\ 3/84 & X_{(3)}=4 \\ 6/84 & X_{(3)}=5 \\ 10/84 & X_{(3)}=6 \\ 15/84 & X_{(3)}=7 \\ 21/84 & X_{(3)}=8 \\ 28/84 & X_{(3)}=9 \\ 0 & \text{el resto de } X_{(3)} \end{cases}$$

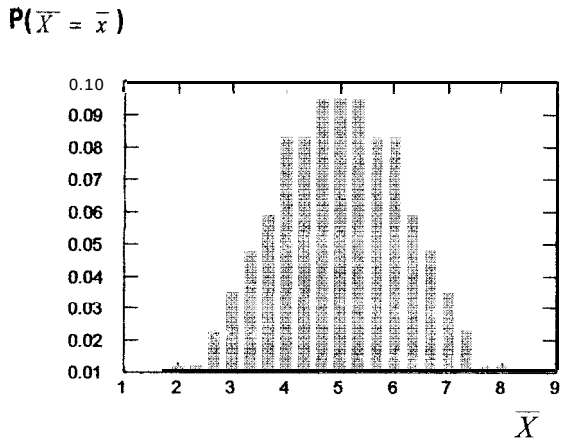
FIGURA 1.3. Histograma de Probabilidades de $X_{(3)}$



Distribución de probabilidades de la media aritmética:

$$P(\bar{X}=\bar{x})=\begin{cases} 1/84 & \bar{X}=2,\ 2.33,\ 7.66,\ 8 \\ 2/84 & \bar{X}=2.66,\ 7.33 \\ 3/84 & \bar{X}=3,\ 7 \\ 4/84 & \bar{X}=3.33,\ 6.66 \\ 5/84 & \bar{X}=3.66,\ 6.33 \\ 7/84 & X=4, 4.33, 5.66,\ 6 \\ 8/84 & \bar{X}=4.66,\ 5,\ 5.33 \\ 0 & \text{el resto de } \bar{X} \end{cases}$$

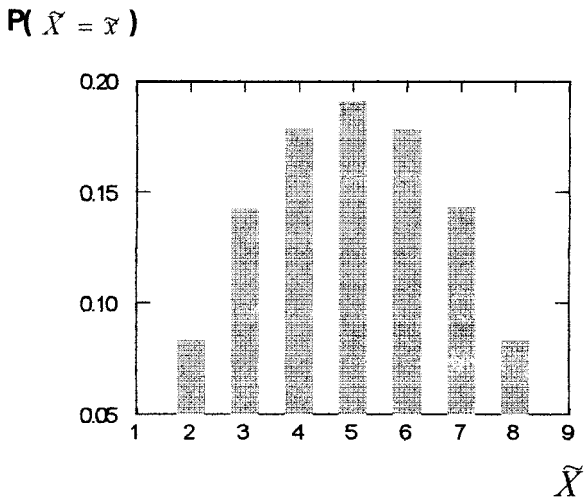
FIGURA 1.4. Histograma de Probabilidades de la Media aritmética



Distribución de probabilidades de la mediana de la muestra:

$$P(\tilde{X}=\tilde{x}) = \begin{cases} 7/84 & \tilde{X}=2, 8 \\ 12/84 & 8=3, 7 \\ 15/84 & \tilde{X}=4, 6 \\ 16/84 & \tilde{X}=5 \\ 0 & \text{el resto de } \tilde{X} \end{cases}$$

FIGURA 1.5. Histograma de Probabilidades de la Mediana.



Como se puede observar en la figura 1.1. el valor mínimo $(X_{(1)})$ con mayor probabilidad de ocurrir es el uno, seguido del dos. En la figura 1.2. notamos que los valores que toma el estadígrafo de orden 2 se distribuyen simetricamente alrededor de su media. Por otra parte, en la figura 1.3. se observa que el valor máximo $(X_{(3)})$ con mayor probabilidad de formar parte de la muestra es el nueve. Nótese que X y $\tilde{X}$ no tienen la misma distribución pero si $\mu_{\bar{X}} = \mu_{\tilde{X}}$.

Además, para el ejemplo que se ha presentado, si al seleccionar la muestra escogemos una con $\overline{X} = 2$, ésta no proporcionaría una buena estimación de la media poblacional (μ=5), sin embargo, cabe señalar que estas muestras tienen menos probabilidad de ocurrir que aquellas que sí dan una "buena" estimación de la media poblacional. En condiciones ideales, se desea que $\overline{X} = \mu$.

En la figura 1.5, se puede observar que la distribución de esta variable coincide con la de la figura 1.2, esto se debe a que la mediana en este ejemplo coincide con el estadígrafo de orden 2, lo cual no hubiese ocurrido si el tamaño de la muestra hubiera sido un número par o mayor a tres.

A continuación se presenta la media, la varianza y la función generadora de momentos para cada una de las variables aleatorias antes mencionadas. (TABLA III)

La función generadora de momentos como se definió anteriormente, viene dada por:

$$M_x[t] = E(e^{tx})$$

TABLA IIL Medias, Varianzas y FGM, N=9,n=3

Variable Aleatoria	Media	Varianza	Función Generadora de Momentos
$X_{(1)}$	2.5	2.251	$1/84(28+21e^{t}+15e^{2t}+10e^{3t}+6e^{4t}+3e^{5t}+e^{6t})$
$X_{(2)}$	5	1.928	$7/84e^{2t}(1+e^{6t})+12/84e^{3t}(1+e^{4t})+15/84e^{4t}(1+e^{2t})+16/84e^{5t}$
$X_{(3)}$	7.5	1.930	$1/84e^{3t}(1+3e^{t}+6e^{2t}+10e^{3t}+15e^{4t}+21e^{5t}+28e^{6t})$
$\overline{X}$	5	1.666	$7/84(e^{4t}+e^{4.33t}+e^{5.66t}+e^{6t})+8/84(e^{4.66t}+e^{5t}+e^{5.33t})$
$\tilde{X}$	5	1.928	$7/84e^{2t}+7/84e^{8t}+12/84e^{3t}+12/84e^{7t}+15/84e^{4t}+15/84e^{6t}$

Como es de esperarse, en la TABLA III se observa que cada variable aleatoria tiene su propia función generadora de momentos la misma que es única para cada una de ellas, así mismo, se puede apreciar que la varianza de la mediana es mayor que la de la media aritmética a pesar de que las dos se encuentran distribuidas alrededor de la media, véanse las figuras 1.4 y 1.5.

A continuación se presentan tres métodos de muestreo que son familiares para quienes realizan este tipo de actividad.

ESPOL
POLITECNICA DEL LITORAL
CIB - ESPOL

1.3 Muestreo Aleatorio Simple.

Antes de describir en que consiste esta técnica, es necesario introducir dos definiciones importantes tales como muestra aleatoria tomada de una población infinita y muestra aleatoria tomada de una población finita.

Muestra aleatoria tomada de una población infinita

Si se tiene una población infinita X, una sucesión de variables aleatorias X_1, X_2, X_3,....X_n constituyen una muestra aleatoria si y sólo si, estas son independientes e indenticamente distribuidas, es decir , que X_i tiene la misma distribución que X, para i=1,2,...,n..

Muestra aleatoria tomada de una población finita

Si se tiene una población finita de tamaño N, el subconjuto de valores X_1, X_2,X_3,...X_n, tomados de X, constituyen una muestra aleatoria si y sólo si, todos y cada uno de los $\binom{N}{n}$ conjuntos de tamaño n tienen igual probabilidad de constituir la muestra.

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

Cuando se tiene una población finita y la muestra que se selecciona se la toma unidad a unidad, sin reposición de estas a la población después de cada selección, se dice que el Muestreo es Aleatorio Simple o h-estrictamente aleatorio.

Nótese entonces que uno de los objetivos del Muestreo aleatorio es tratar de eliminar la predisposición con la que las personas podrían elegir los elementos de una muestra.

Para verificar que bajo estas condiciones cada una de las unidades de la muestra tiene igual probabilidad de ser extraída, se considerará una muestra determinada, es decir, una colección de n unidades especificadas. En la primera extracción la probabilidad de que se seleccione una de éstas n unidades esta dada por $\frac{n}{N}$.

En la segunda extracción la probabilidad de que se seleccione una de las restantes (n-l) unidades es $\frac{n-1}{N-1}$, y así sucesivamente. Por lo tanto la probabilidad de que se seleccionen las n unidades especificadas es:

$$\frac{n}{N} \cdot \frac{(n-1)}{(N-1)} \cdot \frac{(n-2)}{(N-2)} \cdots\cdots \frac{1}{(N-n-1)} = \frac{n!\,(N-n)!}{N!} = \frac{1}{\binom{N}{n}}$$

Es importante señalar que el interés del Muestreo fundamentalmente se centra en tres características de la población:

1. Media Poblacional
2. Total de la Población
3. Proporción de unidades que caen dentro de alguna clase definida.

1.3.1 Propiedades deseables de un estimador.

La precisión de cualquier estimación basada en una muestra depende del método por el cual se efectuó la estimación a partir de los datos de la muestra y del plan de muestreo.

Antes de mencionar las propiedades deseables de un estimador, definamos a qué se denomina ***estimador de un parúmetro.***

El estimador de un parámetro@ es cualquier variable aleatoria $\hat{\theta}_i$ que se exprese en función de la muestra aleatoria y que tenga como objetivo aproximar el valor de θ_i:

$$\hat{\theta}_i : \mathbf{R}^n \rightarrow \mathbf{R}$$

Obsérvese que el estimador ***no es un valor constante,*** sino una variable aleatoria, ya que depende unívocamente de los valores de la muestra y la elección de la muestra es un proceso aleatorio.

Luego de haber definido lo que es un estimador, pasemos a mencionar las características que serían deseables que tuviera un estimador.

Estimador insesgado: Un estimador $\hat{\theta}$ es un estimador insesgado del parámetro θ si y sólo si

$$E(\hat{\theta}) = \theta$$

Estimador Consistente: $\hat{\theta}$ es un estimador consistente del parámetro θ si y sólo si para cada constante c positiva ,

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| \geq c) = 0$$

esto es, $\hat{\theta} \xrightarrow{P} \theta$

Estimador Suficiente: El estimador debería aprovechar toda la información existente en la muestra relevante para la estimación del parámetro θ de la población. Por esto se dice que la estadística $\hat{\theta}$ es un estimador suficiente del

parámetro θ si y sólo si la densidad conjunta o distribución de probabilidad de la muestra aleatoria se puede factor-izar de manera que

$$f(x_1, x_2, \ldots, x_n; \hat{\theta}) = g(\hat{\theta}, \theta).h(x_1, x_2, \ldots x_n)$$

Donde $g(\hat{\theta}, \theta)$ depende solo de $\hat{\theta}$ y θ , y $h(x_1, x_2, \ldots x_n)$ no depende de θ.

Eficiencia: El estimador, al ser una variable aleatoria no puede exigírsele que para una muestra cualquiera tome el valor exacto del parámetro. Sin embargo podemos pedirle que su dispersión (varianza) con respecto al valor central sea tan pequeña como sea posible. Es por esto que si $\hat{\theta}_1$ *y* $\hat{\theta}_2$ son estimadores insesgados del parámetro θ, la eficiencia relativa de los dos estimadores suele compararse a través de sus varianzas, ésto es, $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimaciones insesgadas de un parámetro θ ***y*** $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, ***se*** dice que $\hat{\theta}_1$ es relativamente más eficiente.

Una vez establecidas algunas de las características que uno desea que tenga un estimador, pasemos a definir los estimadores fundamentales para este tipo de muestreo.

1.3.2 Estimaciones.

La media poblacional es estimada por la media aritmética muestral, ésto es

$$\mu = E(X) \qquad \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \hat{\mu}$$

El total poblacional y el de la muestra están definidos por:

$$X = \sum_{i=1}^{N} X_i \qquad \hat{X} = \left(\sum_{i=1}^{n} \frac{x_i}{n} \right) N = N\bar{x}$$

Es necesario definir además de la media y del total, la razón de la muestra con respecto a la población, la misma que se denomina fracción de muestreo y viene dada por $f = \frac{n}{N}$.

La media muestra1 $\bar{x}$ es un estimador insesgado de μ, es decir, $E(\bar{x}) = \mu$, con base en ésto se tiene que 8 = $N\bar{x}$ es también un estimador insesgado del total de la población. Cuando la población es finita de tamaño N la varianza de esta viene dada por:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Mientras que la Cuasivarianza poblacional es denotada por S^2 y definida:

$$S^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N - 1}$$

σ^2 y S^2, siendo medidas de dispersión con respecto a la media tienen estimadores, entre ellos la **varianza** muestra1 s^2 , la misma que viene dada por

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

se puede probar que s^2 es un estimador insesgado de σ^2, esto es

$$E\left[s^2\right] = \sigma^2$$

Consideremos ahora la **varianza** de $\bar{X}$ de una muestra aleatoria simple, se puede probar que

$$Var(\bar{x}) = (1 - f)\frac{S^2}{n}$$

Demostración:

$$n(\overline{x}-\mu) = (x_1-\mu)+(x_2-\mu)+\cdots+(x_n-\mu) \qquad (1)$$

nótese que:

$E(x_1 + x_2 + \cdots + x_n)$ debe ser múltiplo de $x_1+x_2+\ldots\ldots+x_N$

entonces, usando este argumento se tiene que

$$E\left[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2\right]=\frac{n}{N}\left[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2\right]$$

(2)

y que

$$E\left[(x_1-\mu)(x_2-\mu)+(x_1-\mu)(x_3-\mu)+\cdots+(x_{n-1}-\mu)(x_n-\mu)\right]$$

$$=\frac{n(n-1)}{N(N-1)}\left[(x_1-\mu)(x_2-\mu)+(x_1-\mu)(x_3-\mu)+\cdots+(x_{N-1}-\mu)(x_N-\mu)\right]$$

(3)

La suma de productos se extiende sobre todas las parejas de unidades de la muestra y en la población, respectivamente.

De aquí tenemos

$$n^2E(\bar{x}-\mu)^2$$

$$=\frac{n}{N}\left\{(x_1-\mu)^2+\cdots+(x_N-\mu)^2+\frac{2(n-1)}{N-1}[(x_1-\mu)(x_2-\mu)+\cdots+(x_{N-1}-\mu)(x_N-\mu)]\right\}$$

$$n^2\ E(\bar{x}-\mu)^2$$

$$=\frac{n}{N}\left\{\left(1-\frac{n-1}{N-1}\right)\{(x_1-\mu)^2+\ldots+(x_N-\mu)^2\}+\frac{(n-1)}{N-1}[(x_1-\mu)+\ldots+(x_N-\mu)]^2\right\}$$

Dividiendo para n^2 a la ecuación anterior se obtiene,

$$\mathrm{Var}(\bar{x})=E(\bar{x}-\mu)^2=\frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(x_i-\mu)^2=\frac{S^2}{n}\left(1-\frac{n}{N}\right)=\sigma_{\bar{x}}^{\ 2}$$

Lo cual concluye la prueba.

A partir de la **varianza** de la media aritmética se obtienen el error estándar de $\bar{x}$, el mismo que esta dado por,

$$\sigma_{\bar{x}}=\frac{S}{\sqrt{n}}\sqrt{1-ff}$$

Una vez definida la **varianza** y el error de la media muestra1 pasaremos a determinar la **varianza** del estimador del total $\hat{X}$.

$$Var\left(\hat{X}\right) = Var\left(N\bar{x}\right) = N^2 Var\left(\bar{x}\right) = N^2\left(1-f\right)\frac{S^2}{n}$$

Como se menciona anteriormente la cuasivarianza poblacional S^2 no se conoce, y es por esta razón que las varianzas de la media y del total son expresados en términos de un estimador, es decir en términos de s^2, obteniéndose así:

$$\widehat{\mathrm{var}\left(\bar{x}\right)} = \left(1-f\right)\frac{s^2}{n}$$

$$\widehat{\mathrm{var}\left(\hat{X}\right)} = \widehat{\mathrm{var}\left(N\bar{x}\right)} = N^2\,\widehat{\mathrm{var}\left(\bar{x}\right)} = N^2\left(1-f\right)\frac{s^2}{n}$$

Así mismo definamos un intervalo de confianza para el error de estimación partiendo de la ecuación fundamental del muestreo dada por,

$$e = \left|\hat{\theta} - \theta\right| \leq K\sigma_{\hat{\theta}}$$

de ésta tenemos que

$$\left|\bar{x} - \mu\right| \leq Z_{\alpha/2}\hat{\sigma}_{\bar{X}}$$

Cabe acotar que si el tamaño de la muestra es pequeño, entonces se deberá reemplazar en la el $Z_{\alpha/2}$ por $t_{\alpha/2}(n-1)$, donde t es una variable aleatoria que sigue una distribución de Student con (n-l) grados de libertad.

Tamaño de la Muestra

Para establecer el tamaño que debe alcanzar una muestra para la estimación de la media poblacional, hay que tomar en consideración algunos factores importantes como: el error muestra1 admisible e, el nivel de confianza y el tipo de muestreo. Consideremos el error de estimación dado por:

$$\left|\hat{\theta} - \theta\right| = e$$

$$\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq Z_{\alpha/2}$$

$$\Rightarrow \qquad \frac{e}{\sigma_{\hat{\theta}}} \leq Z_{\alpha/2}$$

$$e = Z_{\alpha/2}\sigma_{\hat{\theta}}$$

$\hat{\theta} = \bar{x}$ en muestreo aleatorio simple, por tanto

$$e = Z_{\alpha/2}\frac{S}{\sqrt{n}}\left(\sqrt{1-\frac{n}{n}}\right)$$

$$e^2 = Z^2_{\alpha/2}\frac{S^2}{n}\left(1-\frac{n}{N}\right)$$

$$e^2 = Z^2_{\alpha/2}\frac{S^2}{n} - \frac{Z^2_{\alpha/2}}{N}S^2$$

$$\frac{1}{n} = \frac{e^2 + Z^2_{\alpha/2}S^2/N}{Z^2_{\alpha/2}S^2}$$

De aquí se obtiene el valor n del tamaño de la muestra cuando utilizamos Muestreo Aleatorio Simple:

$$n = \frac{Z^2_{\alpha/2} S^2}{e^2 + \frac{Z^2_{\alpha/2} S^2}{N}}$$

$$n = \frac{Z^2_{\alpha/2} S^2 / e^2}{1 + \frac{Z^2_{\alpha/2} S^2}{e^2 N}}$$

Si se reemplaza a $\frac{Z^2_{\alpha/2} S^2}{e^2}$, por n_o, tenemos

$$n = \frac{n_o}{1 + \frac{n_o}{N}}$$

Ilustración.

Tomemos una localidad que tiene 60 elementos en su Población Económicamente Activa (PEA), en la tabla I se resumen los ingresos que ellos perciben.

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

TABLA IV. Datos de la Ilustración para el Muestreo Aleatorio Simple

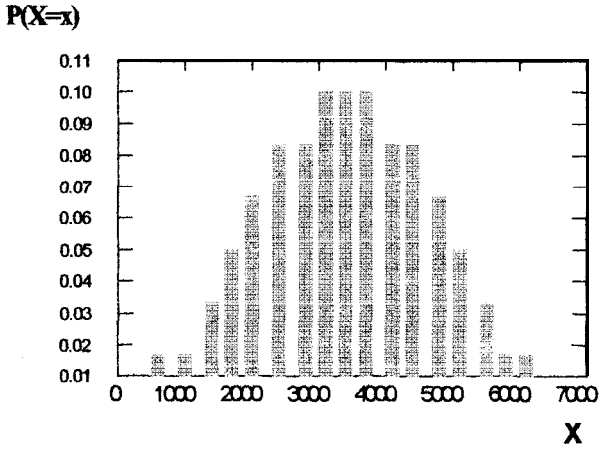
Ingresos	# de habitantes
600.000	1
1'000.000	1
1'400.000	2
1'700.000	3
2'000.000	4
2'400.000	5
2'800.000	5
3'100.000	6
3'400.000	6
3'700.000	6
4'100.000	5
4'400.000	5
4'800.000	4
5'100.000	3
5'500.000	2
5'800.000	1
6'100.000	
Total: 57'900.000	**60**

Se desea estimar el promedio de ingresos (variable X) que tiene esta población, así como también con una confianza del 95% se desea establecer un intervalo para el total de ingresos; para esto se considerará un error no más del 6% del promedio.

A continuación se presenta la distribución de probabilidades que sigue la variable ingresos X de esta población:

$$P(X=x)=\begin{cases} 1/60 & x=600.000,1000.000,6100.000,5800.000 \\ 2/60 & x=1400.000,5500.000 \\ 3/60 & x=1700.000,\ 5100.000 \\ 4/60 & x=2000.000,4800.000 \\ 5/60 & x=2400.000,\ 2800.000,\ 4100.000,\ 4400.000 \\ 6/60 & x=3\ 100.000,34000.000,\ 3\ 700.000 \\ 0 & \text{el resto de } x \end{cases}$$

FIGURA 1.6. Histograma de Probabilidades de los Ingresos de la PEA.



Supongamos que para esto se toma una muestra piloto de tamaño 5, que resulta ser,

{ 1'700.000, 2'000.000, 2'000.000, 1'700.000, 2'400.000}

La media aritmética de la muestra es,

$$\bar{x} = \sum_{i=1}^{5} \frac{x_i}{5}$$

$\bar{x}$ = 1'960.000 sucres

y la varianza estimada

$$s^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})}{4}$$

s^2 = 83.000.000.000 sucres'

Si se desea estimar la media μ de esta población, con una confianza del 95% y un error menor al 6% del promedio, entonces el tamaño de la muestra estará dado por:

$$n = \frac{Z^2_{\alpha/2} S^2 / e^2}{1 + \frac{Z^2_{\alpha/2} S^2}{e^2 N}}$$

con $Z_{\alpha/2}$ =1.96 se obtiene n = 16.4

Ahora pasemos a seleccionar la muestra aleatoria definitiva. Sea la muestra:

{ 1’400.000, 2’000.000, 2’000.000, 3’100.000, 2’400.000, 2’400.000, 3’700.000, 3’700.000, 2’800.000, 2’800.000, 2’800.000, 5’100.000, 1’000.000, 2’000.000, 3’100.000, 4’800.000}

Su media aritmética de está dada por

$$\bar{x} = \sum_{i=1}^{16} \frac{x_i}{16}$$

$\bar{x}$ = 2’818.750 sucres

y la varianza estimada

$$s^2 = \frac{\sum_{i=1}^{16} (x_i - \bar{x})^2}{15}$$

$s^2 = 1.235 \times 10^{12}$. sucres ^{2}

Calculemos ahora la varianza de la media

$$\widehat{\text{var}}(\bar{x}) = (1 - f)\frac{s^2}{n}$$

donde f = 16/60

Reemplazando f, s^2 y n, se obtiene:

$$\widehat{var(\bar{x})} = 56'604.165.805,7$$

Así mismos definamos el límite para el error de estimación como:

$$|\bar{x} - \mu| \leq Z_{\alpha/2}\hat{\sigma}(\bar{x})$$

$E \leq 1.96\ (237916.3) = 466.3\ 15,95$

es decir que un intervalo con el 95% de confianza para $\mu_{\bar{x}}$ es

$\bar{x} \pm 466.315,95$

1.3.3 Estimación para Proporciones

Supongamos que se tiene una variable aleatoria X_i que es Bernoulli, esto es, solo puede tomar los valores de 0 ó 1

$$X_i = \begin{cases} 0 & \text{si no tiene una característica dada} \\ 1 & \text{si tiene la característica} \end{cases}$$

En este caso se define a p como la proporción de elementos de la población que tienen la característica deseada, como A, al número total de elementos que tienen la característica y la misma que estará dada por A= pN , finalmente se define a q como la proporción de elementos que no tienen la característica deseada, esto es 4 = (1-p).

Considerando lo antes mencionado, tenemos que:

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

Luego una estimación de esta proporción viene dada por

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \hat{p}$$

Nótese que aquí el problema de estimar A y p es similar a la estimación del total y de la media de una población en la que todos los valores de X_i son "unos" o bien son "ceros". Para poder determinar la varianza de A y de p es necesario que expresemos a la cuasivarianza S^2, y la varianza muestra1 s^2 en términos de p y $\hat{p}$.

Sabemos que:

$$S^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{\text{M-l}} = \frac{\sum_{i=1}^{N} x_i^{\,2} - N\mu^2}{\text{N - l}}$$

$$S^2 = \frac{N}{N-1}pq$$

De la misma manera, se puede establecer que $s^2 = \hat{\sigma}^2$ y también expresarlo como:

$$s^2 = \frac{n}{n-1}\hat{p}\hat{q}$$

Luego de haber calculado ía **cuasivarianza** S^2 en términos de la proporción, y su estimación s^2, determinemos la **varianza** de la proporción en la muestra, es decir, la **varianza** de $\hat{p}$

$$Var(\hat{p}) = Var(\bar{x}) = (1-f)\frac{S^2}{n}$$

$$Var(\hat{p}) = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$$

Luego la **varianza** de la estimación del número total de unidades que tienen una cierta característica es,

$$Var\left(\hat{A}\right) = \frac{N^2 pq}{n}\left(\frac{N-n}{N-1}\right)$$

Como no se conoce p ni q, se debe realizar una estimación de las varianzas antes mencionadas, ayudándonos para eso de s^2. Por lo tanto tendremos, que una estimación insesgada de la varianza de $\hat{p}$ y de $\hat{A}$ es:

$$\widehat{\mathrm{var}(\hat{p})} = \frac{\hat{p}\hat{q}}{N}\left(\frac{N-n}{n-1}\right)$$

$$\widehat{\mathrm{var}(\hat{A})} = \frac{N(N-n)}{n-1}\hat{p}\hat{q}$$

Finalmente definamos el intervalo para el error de estimación como:

$$\left|\hat{p} - p\right| \leq Z_{\alpha/2}\hat{\sigma}\left(\hat{p}\right)$$

Tamaño de la muestra

En síntesis, el tamaño de la muestra utilizando Muestreo Aleatorio Simple, para determinar la proporción p de elementos de una población de tamaño N que tiene una característica especifica esta dado por:

$$n = \frac{n_o}{1 + \frac{(n_o - 1)}{N}}$$

Ilustración.

Supongamos que una ciudad tiene 15.764 hogares. Se desea conocer cual es el porcentaje de familias que tienen agua potable en sus hogares, cual es el total de familias que carecen de este servicio, el error que se tiene en cada una de las estimaciones, y finalmente se desea establecer un intervalo de confianza del 95% para el número total de hogares que no tienen agua potable.

Si para ese estudio se toma una muestra aleatoria de 350 hogares y en cada uno de ellos se investiga si esa casa es propia o rentada y si tienen o no agua, se obtiene la siguiente información (TABLA V).

TABLA V. Muestreo Aleatorío Simple: Proporciones.

		Si	**No**	
Agua	**Si**	**181**	110	**291**
	No	18	41	**59**
	Total	199	151	**350**

Porcentaje de personas que cuentan con el servicio de agua potable:

$$\hat{p} = \frac{291}{350} = 0{,}83\,14$$

$$e = \sqrt{\frac{\hat{p}\hat{q}}{N}\left(\frac{N-n}{n-1}\right)} = \sqrt{0{,}0003927} = 0{,}01981$$

Estimación del Total:

A=?

$$\hat{p} = \frac{59}{350} = 0{,}1685$$

A=N $\hat{p}$

e = 0,019816

A=(15.764)(0,1685)

A=2.657,36

Intervalo del 95% de confianza:

$$\hat{p} - Z_{\alpha/2}.e < p < \hat{p} + Z_{\alpha/2}.e$$

$$0{,}129 < p < 0{,}207$$

1.4 Muestreo Estratificado

El Muestreo Aleatorio Estratificado es una técnica en la que una población de N individuos se particiona en H subpoblaciones o estratos, cada uno con sus propias características, esto es heterogéneos entre ellos.

Estas subpoblaciones no se traslapan, y en su conjunto abarcan a la población por completo, es decir

$$N = N_1 + N_2 + \dots + N_H ,$$

donde $N_1, N_2, N_3, \dots, N_H$ son los tamaños de cada una de las subpoblaciones.

Una vez que se definen los estratos, se extrae una muestra aleatoria simple de cada uno de ellos, las extracciones deben realizarse en forma independiente en cada uno de los estratos. Los tamaños de las muestras de cada estrato son representados por $n_1, n_2, n_3, \dots n_H$, de tal forma que el tamaño n de la muestra, esta dado por

$$n = n_1 + n_2 + n_3 + \dots + n_H$$

En este tipo de muestreo la media muestral de cada uno de los estratos está dada por:

$$\bar{x}_j = \sum_{i=1}^{n} \frac{x_{hi}}{n_j} \qquad j = 1, 2, 3, \ldots\ldots H$$

Es necesario definir además de la media muestral, la fracción de muestreo, la misma que estará dada en cada uno de los estratos por la relación $\frac{n_h}{N_h}$, y será representada por el símbolo f_h .

1.4.1 Estimaciones

El estimador de la media estratificada $\bar{x}_{est}$ esta definida como

$$\bar{x}_{est} = \sum_{i=1}^{H} \frac{N_h}{N} \bar{x}_h = \hat{\mu}$$

Con base en esto se tiene que 8 = $N\bar{x}_{est}$ es un estimador insesgado del total de la población, por tanto

$$\hat{X} = N_1\bar{x}_1 + N_2\bar{x}_2 + \ldots + N_H\bar{x}_H$$

La varianza de cada estrato viene dada por

$$\sigma_h^2 = \frac{\sum_{i=1}^{N_h}(X_{hi} - \mu_h)^2}{N_h}$$

mientras que la cuasi varianza de cada uno de ellos esta dada por:

$$S_h^2 = \frac{\sum_{i=1}^{N_h}(X_{hi} - \mu_h)^2}{N_h - 1}$$

cuando se toma una muestra aleatoria simple dentro de cada estrato, un estimador insesgado de σ_h^2 es,

$$s_h^2 = \frac{\sum_{i=1}^{n_h}(x_{hi} - \bar{x}_h)^2}{n_h - 1}$$

Consideremos ahora la varianza de la media estratificada ($\bar{x}_{est}$)

$$Var(\bar{x}_{est}) = Var\left(\sum_{i=1}^{H} \frac{N_h}{N} \bar{x}_h\right)$$

$$Var(\bar{x}_{est}) = \sum_{i=1}^{H} W_h^2 Var(\bar{x}_h)$$

$$Var(\bar{x}_{est}) = \sum_{i=1}^{H} W_h^{\,2} \frac{S_h^{\,2}}{n_h}(1 - f_h)$$

donde $W_h = N_h / N$

cuando f_h es pequeña se puede aproximar

$$Var(\bar{x}_{est}) \cong \sum_{i=1}^{H} W_h^{\,2} \frac{S_h^{\,2}}{n_h}$$

Una vez definida la **varianza** de la media estratificada pasaremos a determinar la **varianza** del estimador del total $\hat{X}$.

$$Var(\hat{X}) = \sum_{i=1}^{H} N_h(N_h - n_h) \frac{S_h^{\,2}}{n_h}$$

Como lo que se conoce es el estimador insesgado s_h^2, entonces expresaremos la **varianza** de la media estratificada y la **varianza** de la estimación del total por medio de éste estimador, así tenemos que

$$\widehat{var}(\bar{x}_{en}) = \frac{1}{N^2} \sum_{i=1}^{H} N_k (N_k - n_k) \frac{s_k^{\,2}}{n_k}$$

$$\widehat{var(\hat{X})} = \sum_{i=1}^{H} N_k (N_k - n_k) \frac{s_k^2}{n_k}$$

Así mismo se definen los intervalos de (1-α) 100 % de confianza para la media de la población y para el total de la población:

$$\left|\bar{x}_{est} - \mu\right| \leq Z_{\alpha/2}\hat{\sigma}_{\bar{X}_{est}}$$

$$\left|\hat{X} - N\mu\right| \leq Z_{\alpha/2}\hat{\sigma}_{\hat{X}}$$

$$\bar{x}_{est} \pm Z_{\alpha/2}\sigma_{\bar{x}_{est}}$$

$$N\bar{x}_{est} \pm Z_{\alpha/2}N\sigma_{\bar{x}_{est}}$$

Asignación Optima

En Muestreo Estratificado, los valores de los tamaños de muestra n_h en los respectivos estratos, se los elige de tal forma que minimice $Var(\bar{x}_{est})$ para un costo especifico de tomar una muestra, o, para minimizar el costo para un valor de $Var(\bar{x}_{est})$.

ESPOL
1969
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

Si en cada estrato el costo de observación por unidad es c_h, entonces consideremos la función de costo:

$$C = \sum_{h=1}^{H} n_h c_h$$

Luego, para determinar el tamaño de la muestra óptima para cada estrato es necesario considerar que se debe minimizar la función del costo, de lo que se obtiene

$$n_h = \frac{\frac{N_h}{\sqrt{c_h}} S_h}{\sum_{i=1}^{H} \frac{N_h}{\sqrt{c_h}} S_h} n$$

Como se puede observar, el tamaño de la muestra para cada estrato depende de n, por lo que debemos determinar una asignación óptima para este valor.

Si el costo es fijo, se sustituyen los valores de los n_h óptimos en la función de costo y se resuelve para n, obteniéndose:

$$n = \frac{\sum_{i=1}^{h} S_h^2 W_h^2 / w_h}{\frac{e^2}{Z^2_{\alpha/2}} + \sum_{i=1}^{H} \left(W_h^2 S_h^2\right) / N_h},$$

donde $w_h = \frac{n_h}{n}$

Entonces para plantearnos cuantos elementos deberá tener la muestra es necesario considerar fkdamentalmente dos técnicas: la asignación proporcional y la asignación óptima. A continuación se presenta un ejemplo con el cual se ilustra cada una de estas técnicas.

Ilustración.

Consideremos la siguiente población de 16 elementos dada por:

D={12,2; 11,2; 12,5; 8,3; 8,7; 13,4; 1,7; 12,8; 10,9; 2,3; 6,8; 9,6; 4,3; 10,8; 9,0; 11,4}

con media μ = 8,36875

Si separamos la población en dos estratos, el primero conformado por los elementos menores a 6 y el segundo por los mayores o iguales a 6, y tomamos una muestra aleatoria de tamaño 4, se obtiene:

Estrato 1: {1,2; 1,7; 2,3; 4,3}

Estrato 2: {10,2; 12,5; 13,4; 12,8; 10,9; 10,8; 11,4; 8,3; 8,7; 6,8; 9,6; 11,4; 9,0}

con N_1=4 y N_2= 12

Elementos de la muestra:

Estrato 1: { 1,7; 4,3 }

Estrato 2: { 13,4; 12,8}

con n_1=2 y n_2=2

Se desea estimar el promedio de la población, entonces tenemos que la media de cada estrato es

$$\bar{x}_1 = \sum_{i=1}^{2} \frac{x_{1i}}{n_1} = 3 \qquad \bar{x}_2 = \sum_{i=1}^{2} \frac{x_{2i}}{n_2} = 13,1$$

con una media estratificada igual a

$$\bar{x}_{est} = \sum_{i=1}^{2} \frac{N_h}{N} \bar{x}_h = 8,05$$

y una varianza para cada estrato igual a

$$s_1^2 = \frac{\sum_{i=1}^{2_1}\left(x_{1i} - \bar{x}_1\right)^2}{“1 - 1} = 18{,}12 \qquad s_2^2 = \frac{\sum_{i=1}^{2}\left(x_{2i} - \bar{x}_2\right)^2}{n_2 - 1} = 17{,}061$$

Una vez obtenidas las estimaciones para la varianza y para la media de cada estrato, se procede a calcular la varianza estimada de la media estratificada.

$$\widehat{var}\left(\bar{x}_{est}\right) = \frac{1}{N^2}\sum_{i=1}^{H} N_k\left(N_k - n_k\right)\frac{s_k^2}{n_k}$$

$$\widehat{var}\left(\bar{x}_{est}\right) = 3{,}353437$$

Un intervalo del 95% de confianza para la estimación de la media estaría dado por,

$$\left|\bar{x}_{est} - \mu\right| \leq Z_{\alpha/2}\sigma_{\bar{X}_{est}}$$

$$\bar{x}_{est} \pm 3{,}5892$$

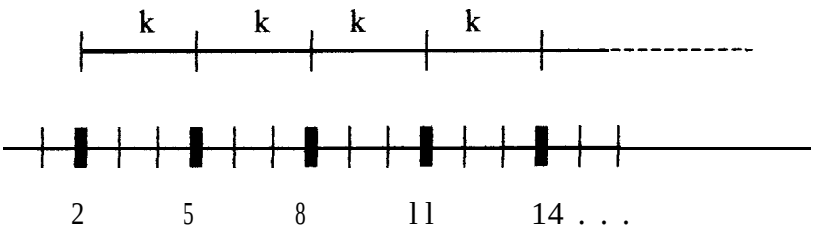
1.5 Muestreo Sistemático.

Cuando se tiene una población de N individuos enumerados desde 1 a N y se toma una muestra de tamaño n, seleccionando la primera unidad al azar de entre las k primeras unidades, luego las subsecuentes a intervalos de k unidades se dice que se esta muestreando sistemáticamente; en este tipo de muestreo se toma a k como el entero más cercano a N/n y generalmente se elige el tamaño de la muestra de la misma manera que en el muestreo aleatorio simple.

Esta técnica, divide la población en n "conglomerados" que consisten en las primeras k unidades, las segundas k unidades, etc.

Supongamos que $k = 3$, y que el primer elemento que conformará la muestra es el de la posición 2, entonces los demás elementos de la muestra serían: $5,8,11$. Esto se puede ver en la figura 1.7 que se muestra a continuación.

FIGURA. 1.7 Composición de las k Muestras



La ventaja que ofrece ésta técnica es que garantiza una distribución espacial de la muestra. De igual manera tiene desventajas, como cuando existe algún tipo de periodicidad en los conglomerados.

En este tipo de muestreo es posible probar que $\bar{x}_{sist} = \frac{\sum_{i=1}^{n} x_i}{n}$ es una estimación insesgada de μ . Por definición la varianza de la población esta dada por:

$$\sigma^2 = \frac{1}{nk}\sum_{i=1}^{N}(X_i - \mu)^2$$

Con base a ésto se tiene que $\hat{X} = N\bar{x}_{sist}$ es un estimador insesgado del total de la población. Consideremos ahora la varianza de $\bar{x}_{sist}$:

$$Var(\bar{x}_{sist}) = \frac{S^2}{n}(1-f)$$

Como lo que se tiene es el estimador s^2, entonces expresaremos la **varianza** de la media sistemática por medio de este estimador, así tenemos que

$$\widehat{var(\bar{x}_{sist})} = \frac{s^2}{n}(1-f)$$

De igual manera definamos los intervalos de confianza para la media de la población como:

$$|\bar{x}_{sist} - \mu| \leq Z_{\alpha/2}\hat{\sigma}(\bar{x}_{sist})$$

$$\bar{x}_{sist} \pm Z_{\alpha/2}\hat{\sigma}(\bar{x}_{sist})$$

A continuación se presenta un ejemplo para ilustrar este tipo de muestreo.

Ilustración

Supongamos que se tiene la siguiente población de tamaño 18 dada por

D = {8, 6, 2, 7, 3, 4, 0, 4, 5, 10, 7, 13, 9, 2, 1, ll, 8, 6}

con media μ = 5,889

Si tomamos el tamaño de la muestra igual a 6, entonces k será igual a 3. Una vez obtenido k, coloquemos los elementos de la población a manera de lista, (ver figura 1 .8)

FIGURA 1.8. Muestras seleccionadas de la Población, N=18.

8	6	2	7	3	4	0	4	5	10	7	13	9	2	1	11	8	6

Si al elegir en forma aleatoria el primer elemento de la muestra se obtiene el número 2 colocado en la tercera posición de la lista, entonces los otros elementos de la muestra serán:

2, 4, 5, 13, 1, 6

Nótese que los elementos son elegidos a intervalos de k elementos, esto es, cada tres, y que en Muestreo Aleatorio Simple, una muestra como 11; 8; 6; 8; 6; 2 puede ser seleccionada, pero en muestreo Sistemático no.

Si calculamos la media sistemática y la varianza, tenemos que

$$\bar{X}_{sist} = \sum_{i=1}^{6} \frac{x_i}{6}$$

$$\bar{x}_{sist} = 5{,}166$$

$$s^2 = \frac{\sum_{i=1}^{6}\left(x_i - \bar{x}_{sist}\right)^2}{5} = 18{,}16$$

La varianza estimada de la media sistemática está dada por,

$$\widehat{var\left(\bar{x}_{sist}\right)} = \frac{s^2}{n}(1-f)$$

$$\widehat{var\left(\bar{x}_{sist}\right)} = 2{,}1186$$

Un intervalo del 95 % de confianza para la estimación de la media estaría dado por,

$$\left|\bar{x}_{sist} - \mu\right| \leq Z_{\alpha/2}\sigma\left(\bar{x}_{sist}\right)$$

$\bar{x}_{sist} \pm 2{,}8528$

ESPOL
1959
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

CAPITULO II

COMPARACIÓN DE LAS PROPIEDADES ANALÍTICAS DEL MUESTREO ALEATORIO SIMPLE, ESTRATIFICADO Y SISTEMÁTICO.

2.1 Generalidades.

En este capítulo se realizará un análisis de algunas propiedades que poseen los estimadores que se obtienen con los diferentes tipos de muestreo antes mencionados; de la misma manera se efectuarán comparaciones entre ellos para de alguna forma determinar cual nos permite obtener una mejor estimación de un parámetro de interés y en que condiciones esto ocurre.

El estudio se enfocará en dos estimadores, sobre los cuales frecuentemente se centra el interés del muestreo, estos son, la media y el total; para el efecto hemos utilizado la simulación estocastica, la cual es una técnica numérica que permite generar variables aleatorias uniformes o no.

Antes de proseguir con el desarrollo de este tema es necesario introducir el concepto de precisión, ya que ésta será una de las propiedades que analizaremos en cada uno de los estimadores obtenidos con los diferentes tipos de muestreo.

- **Precisión:** es la concentración de las estimaciones con respecto a su valor medio.

Esta será medida a través de la **varianza** de estimador, es decir $E\left((\hat{\theta} - E(\hat{\theta}))\right) = Var(\hat{\theta})$ de tal modo que la medida de la precisión estará defmida por:

$$\frac{1}{Var(\hat{\theta})}$$

De esta relación se puede deducir que a medida que la **varianza** aumenta, la precisión disminuye y viceversa. Cabe anotar que otro punto importante es la ***eficiencia relativa*** de un estimador, la misma que como se mencionó en el capítulo anterior se suele comparar a través de sus varianzas. A continuación, se realizará el estudio para cada una de las técnicas de muestreo.

2.2 Comparación de los Resultados obtenidos con las Diferentes Técnicas de Muestreo.

Para este estudio se generaron 4.200 valores de una variable aleatoria N(2,8), 3.000 valores de una N(0,30) y 2.800 valores de una variable aleatoria N(1,6), los mismos que pasaron a conformar el marco muestral.

Una vez definido el marco, el mismo que a su vez paso a constituir la población,, calculamos su media μ y su varianza σ^2, de lo cual obtuvimos

$$\mu = 1{,}141545 \text{ y } \sigma^2 = 14{,}7998$$

Además de la media y de la varianza, calculamos el total de elementos que eran menores o igual a 1,14, obteniéndose

$$A = 4.730$$

Nótese que todos estos valores (parámetros poblacionales) serán estimados a través del muestreo aleatorio simple, del estratificado, y del sistemático, para luego realizar las comparaciones respectivas entre los diferentes métodos.

Del marco muestral tomamos 100 muestras aleatorias de tamaño n=437 con cada uno de los tipos de muestreo; cabe señalar que el tamaño *n* fue determinado a través de una muestra piloto en la cual se estimó la media y la varianza, para luego calcular el tamaño de muestra adecuado con un error menor al 3% del promedio utilizando muestreo aleatorio simple.

Nótese que para el muestreo estratificado utilizamos el mismo tamaño de muestra obtenido para el aleatorio simple e hicimos, una asignación proporcional.

De cada una de las muestras obtenidas con estas técnicas, estimamos la media μ y el total A, recuerdese que μ = 1.1415 y el total es A = 4.730. A continuación se presentan los resultados obtenidos.

TABLA VI. Cuadro Comparativo del Estimador de la media poblacional, n= 437

	Muestreo Aleatorio Simple	**Muestreo Estratificado**	**Muestreo Sistemático**
Número de Casos	100	100	100
Mínimo	0,37400	0,74947	0,69628
Máximo	1,62003	1,55017	1,63313
Mediana	1,13979	1,14295	1,13124
Media	1,10800	1,14047	1,14570
Desviación Estándar	0,20737	0,16847	0,20657
Varianza	0,043 00	0,02838	0,04267

De la TABLA VI podemos observar que el muestreo estratificado proporciona una estimación de la media mucho más próxima al valor real que las otras dos técnicas, en éste, la media estimada es 1,14047 ± 0,16847, en el aleatorio simple es 1,10800 ± 0,20737 y en el sistemático es 1,14570 ± 0,20657 .

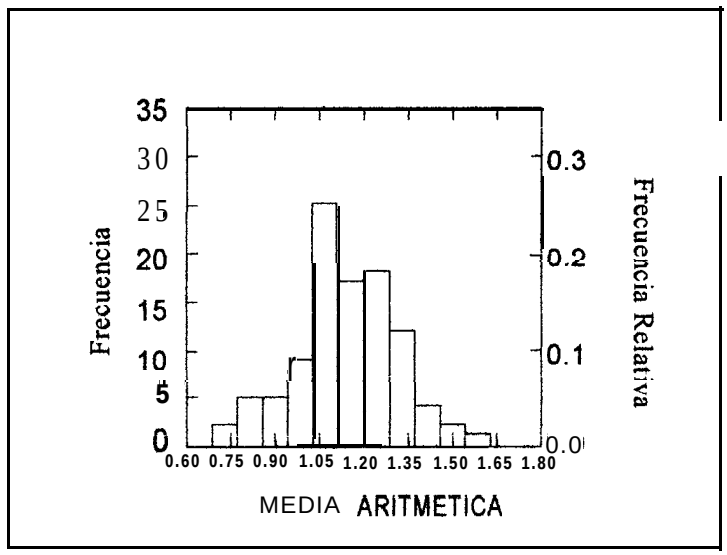
TABLA VII. Cuadro Comparativo del estimador del Total poblacional, n=437

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
Número de Casos	100	100	100
Mínimo	4.18764302	4.120,43372	3.958,81007
Máximo	5.37757437	5.447,56356	5.33 1,80778
Mediana	4.73684210	4.761,53444	4.691,07552
Media	4.73638444	4.728,73242	4.700,45767
Desviación Estándar	23 5,75460	233,13462	280,207 15
Varianza	55.580,23034	54.351,74926	78.516,04874

En la TABLA VII observamos que también el muestreo estratificado proporciona una estimación del total más próxima al valor real (4.73'0) que las otras dos técnicas, nótese que en éste, la media del total es 4.728,73 ± 233,13, en el aleatorio simple es 4.736,38 ± 235,75 y en el sistemático es 4.700 ± 280,20.

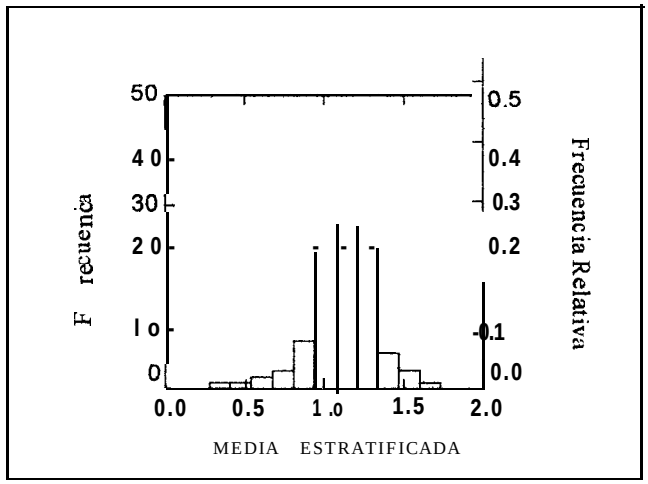
En la figura 2.1, 2.2 y 2.3 se presentan los histogramas de probabilidades para cada una de estas variables.

FIGURA 2.1. Histograma de Frecuencia para la Media Aritmética (Muestreo Aleatorio Simple, n=437)



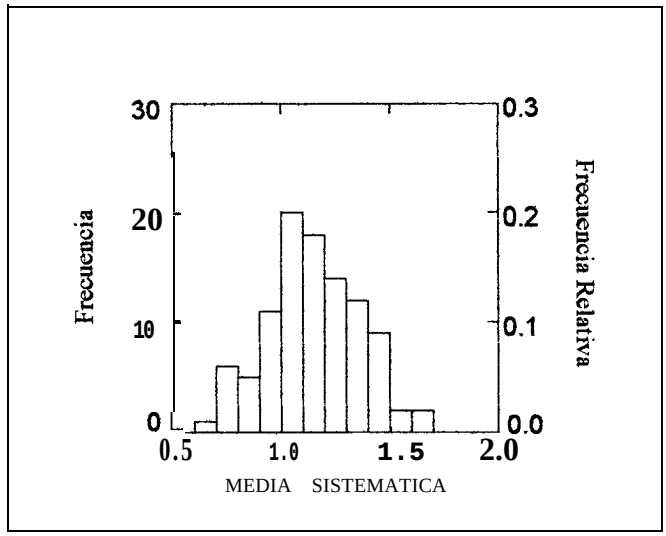
Al analizar esta figura podemos observar que la media aritmética se distribuye alrededor de su media, esto es *1,10800* con una desviación de *0,20737* (1,108 ± 0,20737).

FIGURA 2.2. Histograma de Frecuencia para la Media Estratificada (Muestreo Estratificado, n=437)



En la figura 2.2 observamos que la medía estratificada se encuentra distribuida alrededor de su valor medio, es decir 1,14047, con una desviación de 0,16847; si comparamos la figura 2.1 con la 2.2, observamos que en la primera las observaciones se encuentran más dispersas, mientras que en la figura 2.2 están más próximas a su valor medio $(1{,}14047 \pm 0{,}16847)$.

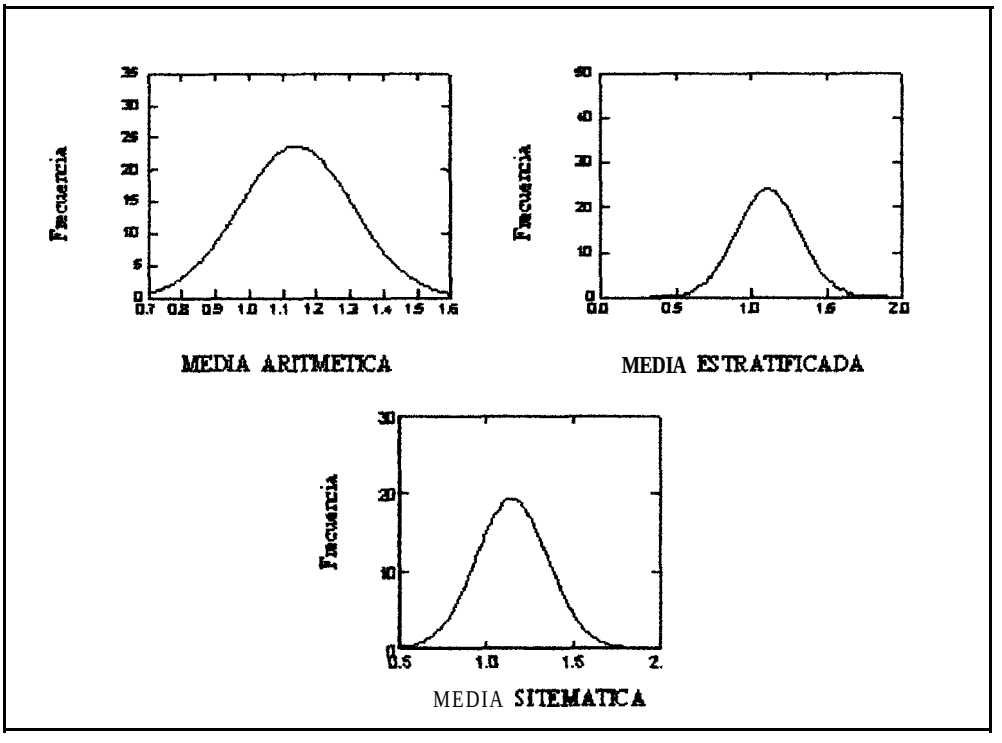
FIGURA 2.3. Histograma de Frecuencia para la Media Sistemática (Muestreo Sistemático, n=437)



En la figura 2.3 la media sistemática se encuentra distribuida alrededor de su valor central, esto es 1,14570 ± 0,20657 ; si la comparamos con las dos anteriores vemos que sus observaciones se encuentran menos dispersas que las de la media aritmética del muestreo aleatorio, pero más que los de la media estratificada.

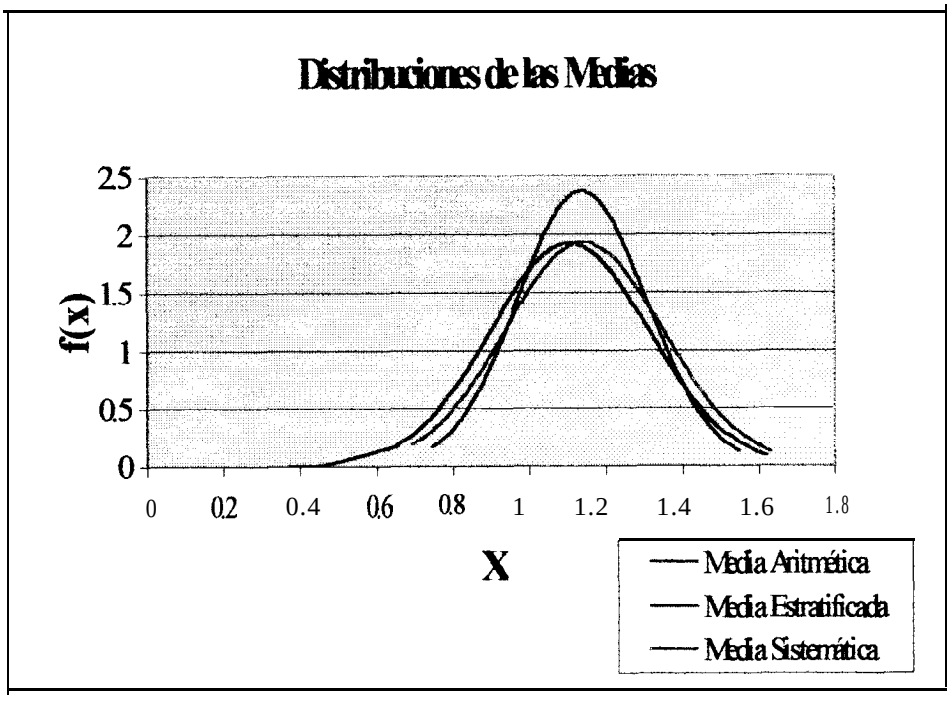
En la figura 2.4 se presentan las distribuciones de cada una de las medias de una manera gráfica.

FIGURA 2.4. Varianzas de los Estimadores



Como se puede observar en la figura 2.4, las medias tienen distribuciones diferentes, cada una de ellas presenta una dispersión diferente de las observaciones. En la figura 2.5 podemos comparar las varianzas que tienen cada uno de estos estimadores con los diferentes métodos de muestreo, donde podemos damos cuenta que el estimador de la media obtenido con el muestreo estratificado ($\bar{x}_{est}$) es el que menor varianza posee, seguido del obtenido con el muestreo sistemático ($\bar{x}_{sist}$).

FIGURA 2.5. Distribuciones de las Medias, n=437



Según la figura 2.5, el estimador con menor varianza es el del muestreo estratificado, seguido del sistemático. Si vamos a la TABLA VI, y calculamos la precisión que tiene cada uno de estos estimadores, obtenemos los siguientes resultados.

Muestreo Aleatorio Simple

$$\text{Precisión } \sigma_{\bar{x}} = \frac{1}{0{,}04300} = 23{,}25$$

ESPOL
1958
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

***Muestreo** Estratificado*

$$\text{Precisión}\left(\bar{x}_{est}\right) = \frac{1}{0{,}02838} = 35{,}23$$

***Muestreo** Sistemático*

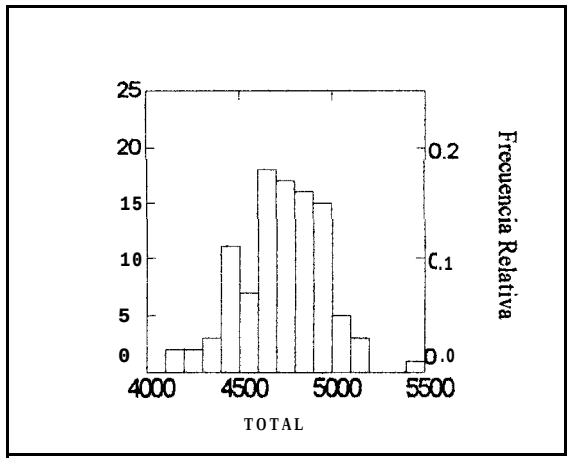
$$\text{Precisión}\left(\bar{x}_{sist}\right) = \frac{1}{0{,}04267} = 23{,}44$$

De aquí podemos establecer que el método con mayor precisión es el estratificado, seguido del sistemático, aunque con poca diferencia significativa del aleatorio simple.

Con respecto a la ***eficiencia*** de estos estimadores observamos que $\boldsymbol{Va+)} > Var\left(\bar{x}_{est}\right)$, ***y*** que $Var\left(\bar{x}\right) > Var\left(\bar{x}_{sist}\right)$, por lo que, bajo estas condiciones $\bar{x}_{est}$ y $\bar{x}_{sist}$ son más eficientes que $\bar{x}$.

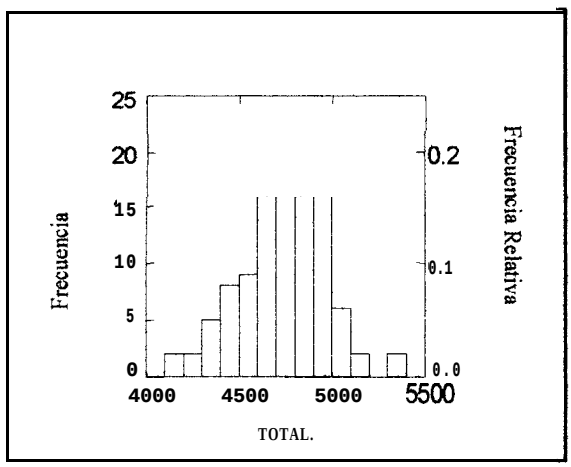
A continuación analizaremos las distribuciones de los totales obtenidos con cada una de estas técnicas.

FIGURA 2.6. Histograma de Frecuencia para el Total (Muestreo Aleatorio Simple,n=437)



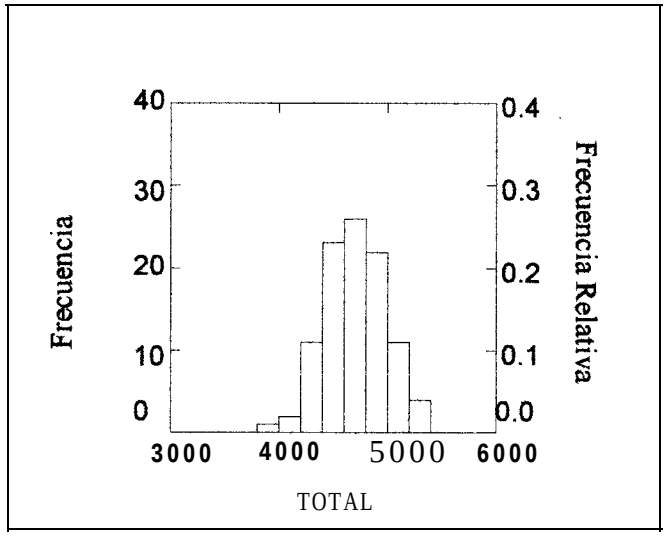
La figura 2.6 muestra la distribución del total obtenida con el muestreo aleatorio simple, como podemos observar se distribuye alrededor de $4.736{,}38 \pm 235{,}75$.

FIGURA 2.7. Histograma de Frecuencia para el Total (Muestreo Estratificado, n=437)



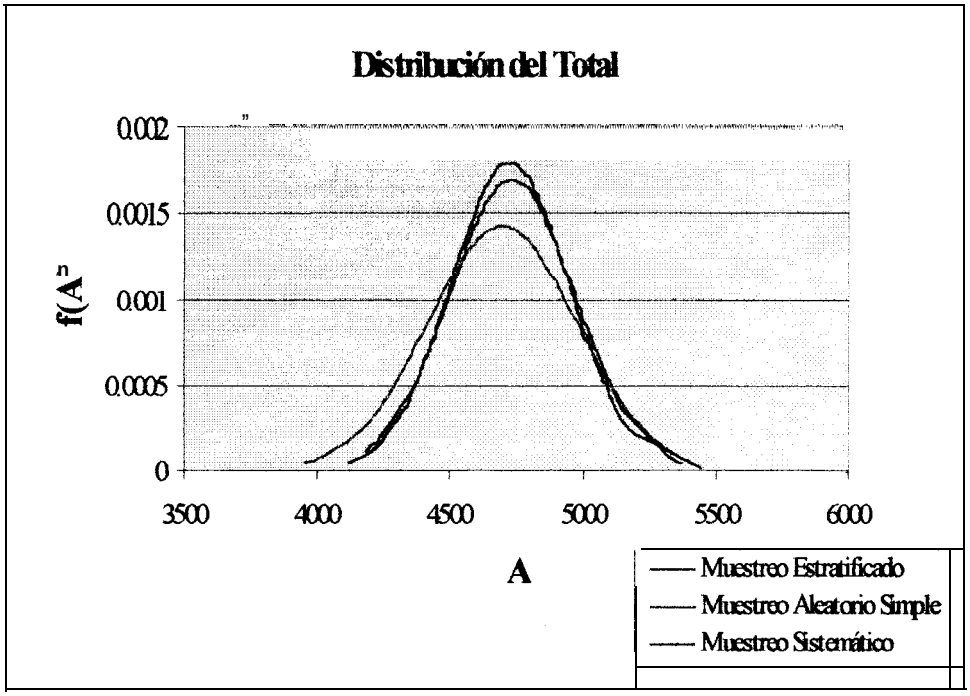
Como podemos observar en esta figura, la media del total se encuentra en 4.728,73 ± 233,1346, si comparamos la dispersión de estas observaciones, con las obtenidas con el muestreo aleatorio simple, podemos observar que estos tiene una menor varianza.

FIGURA 2.8. Histograma de Frecuencia para el Total (Muestreo Sistemático, n=437)



Como podemos observar en la figura 2.8, la media del total se encuentra en 4.700, si comparamos la dispersión de estas observaciones, con las obtenidas con el muestreo aleatorio simple, podemos observar de la TABLA VII que estos tiene una mayor dispersión. Al analizar los valores de las varianzas se obtuvo el siguiente gráfico.

FIGURA 2.9 Distribución del Total, n-437



Como se puede observar en la figura 2.9, el total tiene una distribución diferente con cada una de las técnicas de muestreo. En la figura podemos comparar las varianzas que tienen cada uno de estos estimadores con los diferentes métodos de muestreo, así podemos darnos cuenta que el estimador del total del muestreo estratificado proporciona un menor varianza, seguido del estimador obtenido con el muestreo aleatorio simple.

Si observamos la TABLA VII, y calculamos la precisión que tiene cada uno de estos estimadores, obtenemos los siguientes resultados:

Muestreo Aleatorio Simple

$$\text{Precisión}(A) = \frac{1}{5.5580,23} = 0,0000\ 1799$$

Muestreo Estratificado

$$\text{Precisión}(A) = \frac{1}{5.435,74} = 0,0001839$$

Muestreo Sistemático

$$\text{Precisión}(A) = \frac{1}{7.85\ 16,048} = 0,0000127$$

De aquí podemos observar que el método con mayor precisión es el estratificado, seguido del aleatorio simple.

CAPITULO III

VERIFICACION DE LAS PROPIEDADES ANALITICAS DE LOS DIFERENTES TIPOS DE MUESTREO UTILIZANDO SIMULACION ESTOCASTICA.

En el presente capítulo trabajaremos con poblaciones que a pesar de ser finitas son lo suficientemente grandes como para ser tratadas como de tamaño infinito; considerando este hecho, el estudio que realizaremos estará enfocado en el análisis de la distribución de probabilidad de la media aritmética, así como también en las variables aleatorias total y proporción; se determinará además la "cobertura" que tienen los intervalos de confianza construidos para los parámetros poblacionales.

Antes de continuar con el desarrollo de este capítulo, es necesario recordar el Teorema del Límite Central, uno de los teoremas más importantes de la Estadística, que concierne a la distribución limitante de la media estandarizada de n variables aleatorias independientes e idénticamente distribuidas, cuando $n \to \infty$; así como también presentar la noción que sustenta la construcción de los intervalos de confianza.

Teorema.-

Si $X_1, X_2, X_3, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas de una población cualquiera con media μ y varianza σ^2, sea $\overline{X} = \dfrac{X_1 + X_2 + X_3 + \ . \ . \ . \ + X_n}{n}$, entonces

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{L} Z, \ Z \sim N(0,1),$$

esto es, $\dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ converge en ley a una variable aleatoria normal estándar.

Lo anteriormente expuesto es equivalente a:

$$\overline{X} \xrightarrow{L} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Intervalos de Confianza.-

Si $\hat{\theta}$ es un estimador de θ y se puede determinar dos cantidades que dependan de $\hat{\theta}$, por ejemplo $\hat{\theta}_1$ y $\hat{\theta}_2$ tales que:

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha,$$

para algún $0 \leq \alpha \leq 1$.

Se dice entonces que $\hat{\theta}_1$ y $\hat{\theta}_2$ determinan un intervalo con $(1 - \alpha)$ 100% de confianza para θ. Las cantidades $\hat{\theta}_1$ y $\hat{\theta}_2$ toman el nombre de límite inferior y límite superior del intervalo respectivamente.

Una vez que hemos revisado estos conceptos, se pasará a realizar el estudio para cada una de las técnicas de muestreo. En la primera parte de éste capítulo compararemos los resultados obtenidos con el Muestreo Aleatorio Simple vs. los obtenidos utilizando el Muestreo Sistemático, y en la segunda parte, realizaremos un análisis de las tres técnicas en conjunto, es decir, el aleatorio simple, el estratificado y el sistemático.

3.1 Análisis de los Resultados obtenidos con el Muestreo Aleatorio Simple y con el Muestreo Sistemático.

Para esta parte del estudio se generaron 10.000 valores de una variable aleatoria N(2, 10), los mismos que forman el marco muestral. Después se tomaron 100 muestras, cada una de tamaño 30, analizando la distribución de la media aritmética, la varianza, el total y la proporción, la misma que se define como la proporción de elementos que son menores o iguales a 2.

Para el análisis de la distribución de $\bar{X}$ se utilizará la prueba no paramétrica de Kolmogorov – Smirnov (K- S), la cual es una prueba de bondad de ajuste en la que se establecen dos hipótesis, la hipótesis nula en la que se especifica que distribución tiene la variable aleatoria de interés y es representada por H_0 y la alterna representada por H_1 que niega H_0 .

La prueba K-S define una medida estadística D que se basa en la diferencia máxima entre la función de distribución $F(x)$ que se propone en la hipótesis nula, y la función de distribución empírica $\hat{F}_n(x)$, tal que

$$D = \underset{x}{\text{Max}} \left| F(x) - \hat{F}_n(x) \right|$$

Se acepta H_0 si $D < d_{\alpha,n}$, siendo n el tamaño de la muestra, α el nivel de significancia, y $\hat{F}_n(x)$ igual a

$$\hat{F}_n(x) = \begin{cases} \dfrac{i-1}{n} & \text{si } X_{(i-1)} \le X < X_i \quad i=1,2,\ldots,n \\ 1 & \text{si } X \ge X_n \end{cases}$$

$d_{\alpha.n}$ es un valor tabulado.

En la TABLA VIII se presentan los resultados obtenidos con cada uno de los tipos de muestreo:

TABLA VIII. Comparación de los Resultados para el Muestreo Aleatorio Simple y el Muestreo Sistemático (n = 30)

Muestreo Aleatorio Simple		Muestreo Sistemático	
MEDIA		MEDIA	
Número de Casos	100	Número de Casos	100
Mínimo	0,30498	Mínimo	0,58822
Máximo	3,64387	Máximo	3,37266
Mediana	1,97833	Mediana	1,95839
Media	1,98253	Media	1,91495
Desviación Estándar	0,59630	Desviación Estándar	0,578 ll
Varianza	0,35558	Varianza	0,33421

TOTAL		TOTAL	
Número de Casos	100	Número de Casos	100
Mínimo	2.666,7	Mínimo	2.666,70
Máximo	7.666,7	Máximo	7. 000,00
Mediana	4.700,0	Mediana	5.000,00
Media	4.710,7	Media	5.083,30
Desviación Estándar	969,5	Desviación Estándar	834,7
Varianza	939930,2	Varianza	696724,0
PROPORCION		PROPORCION	
Número de Casos	100	Número de Casos	100
Mínimo	0,26667	Mínimo	0,26667
Máximo	0,76667	Máximo	0,70000
Mediana	0,47000	Mediana	0,50000
Media	0,47107	Media	0,50833
Desviación Estándar	0,09695	Desviación Estándar	0,08347
Varianza	0,00940	Varianza	0,00697

Como se puede observar en la TABLA VIII con el muestreo sistemático se obtiene una media para $\bar{x}_{sist}$ igual a 1,914 15 $\pm$ 0,578 11, mientras que con el muestreo aleatorio simple, la media de $\bar{x}$ se encuentra en 1,9825 $\pm$ 0,5963 .

Nótese que los valores a los cuales debería converger la media y la varianza de $\bar{x}$ son 2 y 0,3333 respectivamente, por lo que observamos que con el muestreo sistemático la varianza de $\bar{x}_{sist}$ se aproxima más al valor que propone el Teorema del Límite Central, aunque la estimación de la media no es "buena" al compararla

con la obtenida por medio del Muestreo Aleatorio Simple. Con respecto a la estimación del total y de la proporción, si comparamos el valor real del total de la población que es 5.027, con cada una de las estimaciones obtenidas, podemos observar que con el muestreo sistemático se obtiene una mejor estimación de la proporción (ver TABLA VIII).

Para probar si la media aritmética y la media "sistemática" se distribuían según el teorema del límite central, realizamos la prueba K-S con las hipótesis

H_0: la media aritmética tiene una distribución N(2 , 0,333)

VS.

H_1: $\neg H_0$.

TABLA IX. Prueba K-S para las medias $\bar{x}$ y $\bar{x}_{sist}$

Muestreo Aleatorio Simple			
Prueba de Kolmogorov -Smimov usando una distribución Normal (2,00 ,0,33)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,04149	0,72533
Muestreo Sistemático			
Prueba de Kolmogorov -Smimov usando una distribución Normal (2,00, 0,33)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,09504	0,67457

Como se puede observar en la TABLA IX, el valor p de la prueba en ambos tipos de muestreo nos dice que existe evidencia estadística para no rechazar la hipótesis nula H_0.

En las figuras 3.1 y 3.2 se presentan los histogramas de frecuencia para la media, con cada una de las técnicas:

FIGURA 3.1. Histograma de Frecuencia para la Media Aritmética (Muestreo Aleatorio Simple,n=30)

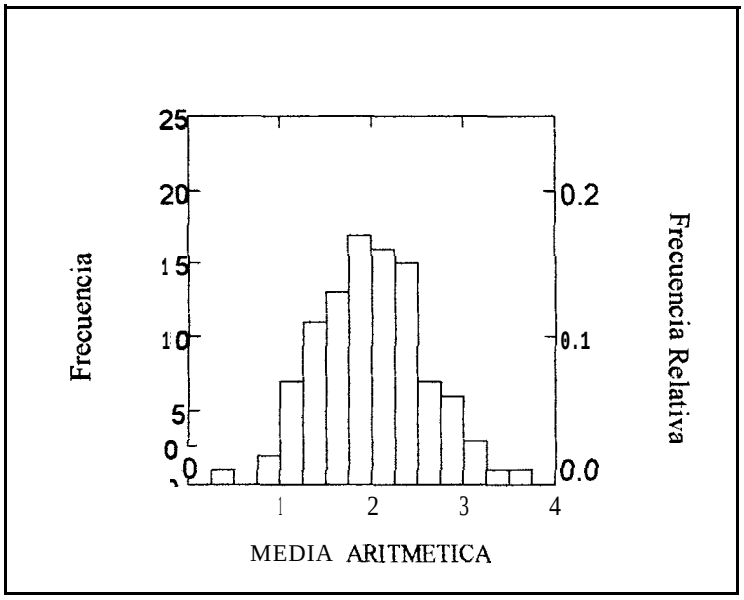
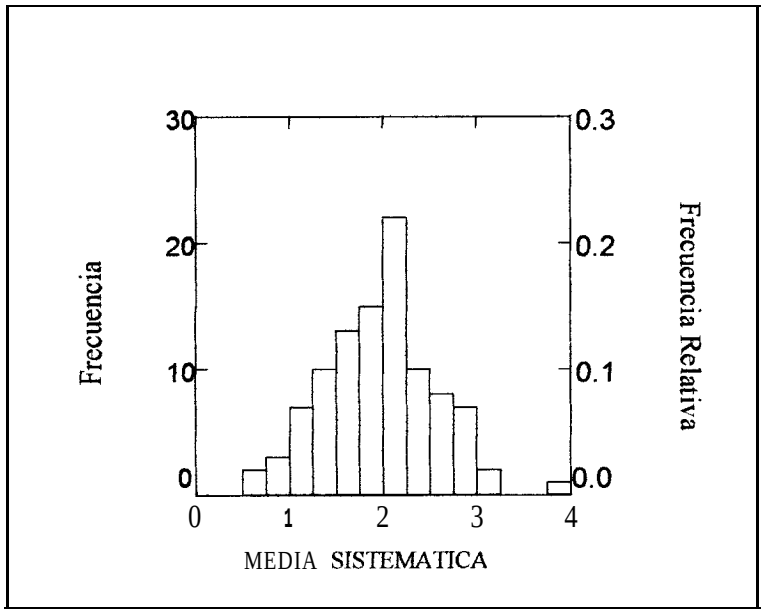


FIGURA 3.2. Histograma de Frecuencia para la Media Sistemática

(Muestreo Sistemático, n=30)



Al analizar tanto la figura 3.1 como la 3.2, se puede observar que ambas variables, tanto la media aritmética del muestreo aleatorio simple ($\bar{x}$) como la media sistemática ($\bar{x}_{sist}$), se distribuyen alrededor de 2, pero cada una con **varianzas** diferentes. Esto se presenta en una forma gráfica en las figuras 3.3 y 3.4 que se presentan a continuación.

FIGURA 3.3. Varianza de la Media Aritmética - N(2,10), n = 30

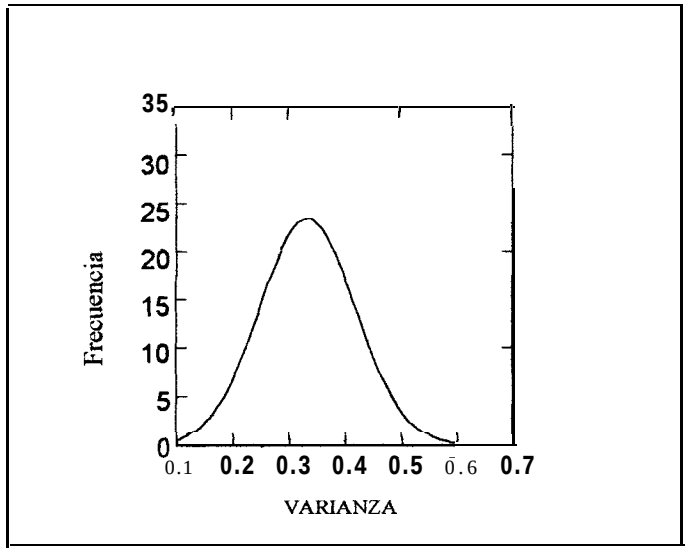
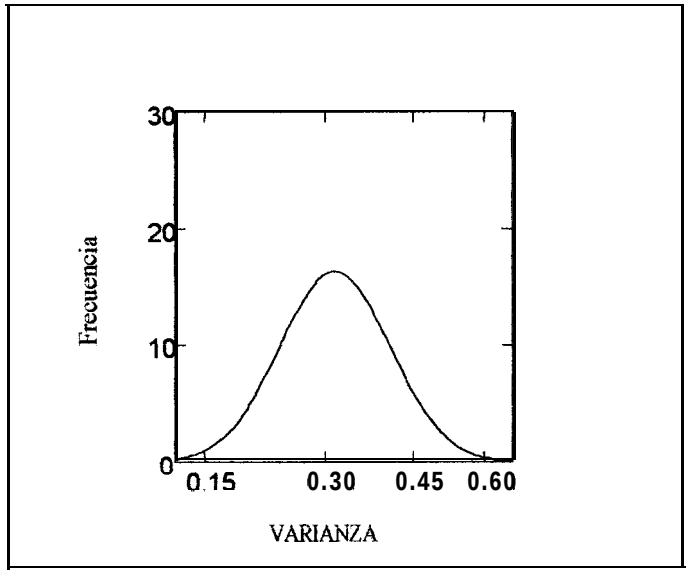


FIGURA 3.4. Varianza de la Media Sistemática - N(2,10), n = 30



Como se mencionó anteriormente en la simulación resulta que la varianza de la media sistemática es menor que la varianza de la media aritmética del Muestreo Aleatorio Simple. (ver figuras 3.3. y 3.4.)

En la figura 3.5 y 3.6 se presentan los gráficos de las distribuciones para el total del muestreo aleatorio simple y del sistemático respectivamente.

FIGURA 3.5. Histograma de Frecuencia para el Total

(Muestreo Aleatorio Simple, n=30)

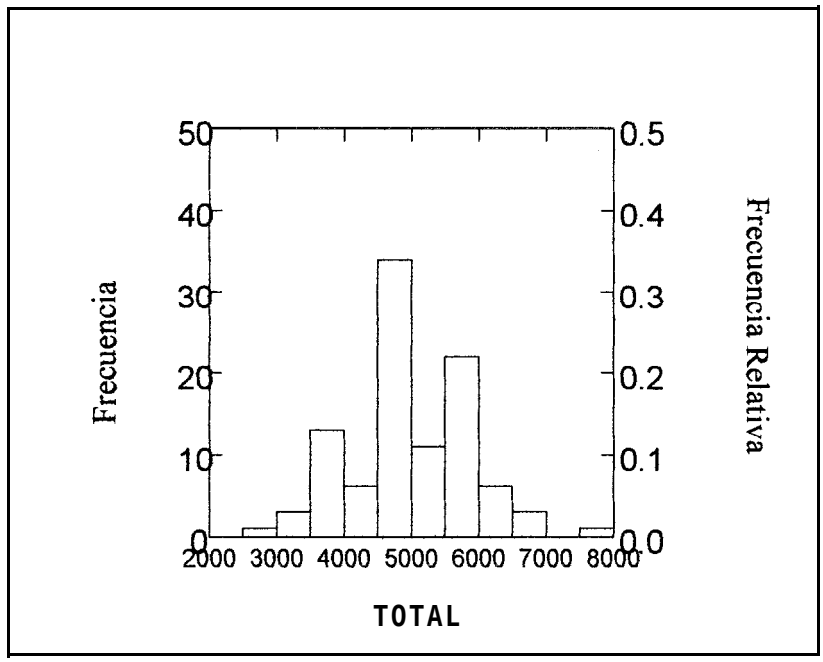
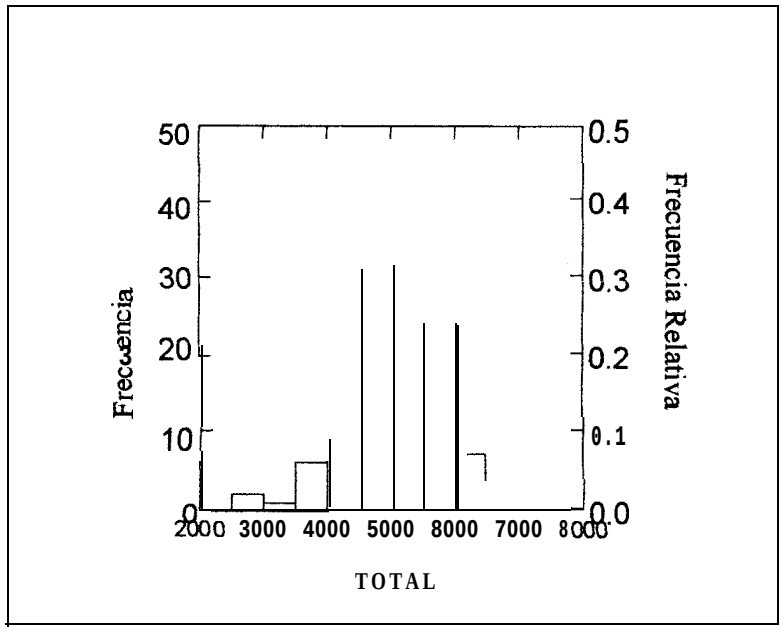


FIGURA 3.6. Histograma de Frecuencia para el Total

(Muestreo Sistemático, n=30)



Como se observa en las figuras 3.5 y 3.6, la varianza del total en el muestreo aleatorio simple es mayor que la del muestreo sistemático.

En la figura 3.7 y 3.8 se presentan las distribuciones para la proporción, del muestreo aleatorio simple y del sistemático respectivamente. Como se mencionó anteriormente, la varianza de la media sistemática es menor a la varianza de la media aritmética.

FIGURA 3.7. Histograma de Frecuencia para la Proporción

(Muestreo Aleatorio Simple, n=30)

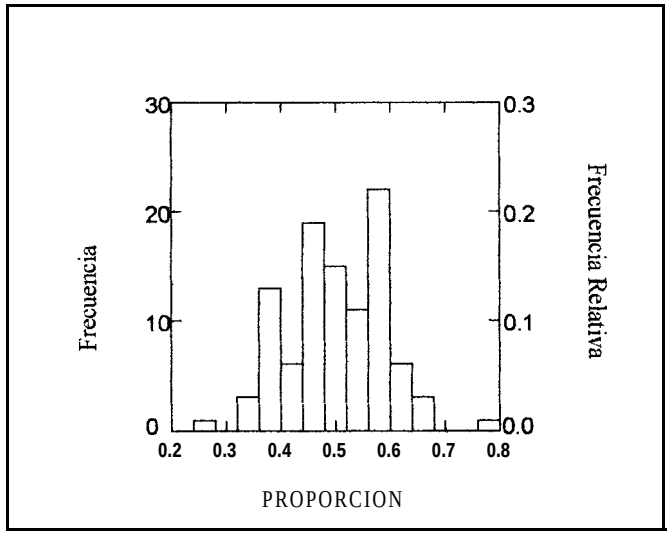
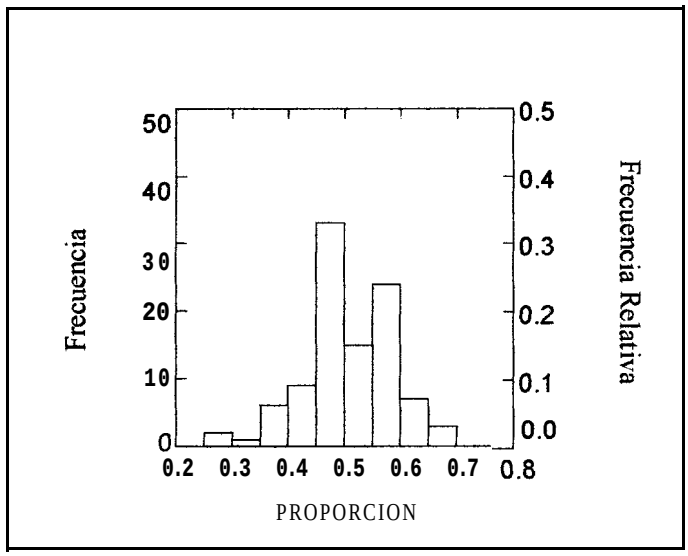


FIGURA 3.8. Histograma de Frecuencia para la Proporción

(Muestreo Sistemático, n=30)



Al analizar las figuras 3.7 y 3.8 se puede ver que la distribución que tiene la proporción en el Muestreo Aleatorio Simple es diferente a la del Muestreo Sistemático; la varianza en el primero es mayor que en el segundo, esto se puede observarse también en la TABLA VIII.

Para analizar la cobertura de los intervalos de confianza correspondientes a cada técnica, construimos intervalos del 95% de confianza para la media aritmética y para la media sistemática; de los cien intervalos construidos, encontramos tanto en el muestreo aleatorio simple como en el sistemático, que seis de los intervalos generados no contenían al valor de la media verdadera, es decir al 2. (ver ANEXO 1).

Nótese que desde el punto de vista teórico, al construir k intervalos con el 95% de confianza, por lo menos el 95 % de ellos deberían contener (cubrir) al parámetro estimado, sin embargo al analizar las tablas del ANEXO 1, podemos observar que esto no se cumple exactamente para n = 30. Así se encuentra que noventa y cuatro de los cien intervalos contienen al valor promedio verdadero, en tanto que la teoría dice que deberían incluirlo al menos noventa y cinco de los cien intervalos. Tal vez esto ocurre debido a que se está considerando un tamaño de muestra que ésta al borde del tamaño finito sugerido para que se aplique el Teorema del Límite Central.

Al analizar los intervalos de confianza de cada una de las técnicas y calcular la longitud de sus intervalos L, definida como

$$L = LCS - LCI,$$

donde LCS es el límite de confianza superior y LCI es el límite de confianza inferior, encontramos que la longitud promedio de los intervalos en el muestreo aleatorio simple es de 2,2497, mientras que en el muestreo sistemático es de 2,2487.

Si consideramos que una propiedad deseable de los intervalos de confianza es que su longitud sea lo más corta posible, podriamos concluir que el muestreo sistemático es mejor el aleatorio simple.

3.2 Análisis de los Resultados obtenidos con el Muestreo Aleatorio Simple, con el Muestreo Sistemático y con el Estratificado.

Para realizar éste análisis se generaron 3.333 valores de una variable aleatoria N(0, 30), y de una variable N(2, 10), y 3.334 valores de una Normal (5, , Estos valores conformaron el marco muestral. Nótese que para esta segunda

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

parte, se muestreo de una población que contiene a su vez subpoblaciones, cada una con diferente media y varianza.

Una vez construido el marco, con cada una de las técnicas tomamos 100 muestras cada una de ellas de tamaño 100, y realizamos el mismo análisis que en la sección anterior, es decir, estimamos el total, la proporción, calculamos la varianza de la media x, x_{est}, y $\bar{x}_{sist}$, etc. La proporción en este caso es definida como la proporción de elementos menores o iguales a 2,3 1.

En la TABLA X se presentan los resultados obtenidos con cada una de las técnicas de muestreo y como podemos observar la técnica que nos proporciona una menor varianza para el estimador $\bar{x}$, es el Muestreo Estratificado, seguido del Sistemático; con el primero obtenemos una media para $\bar{x}_{est}$ igual a 2,29863 ± 0,39552; para la segunda técnica la media para $\bar{x}_{sist}$ es igual a 2,36033 ± 0,404, mientras que con el muestreo aleatorio simple, la media de $\bar{x}$ se encuentra en 2,36805 ± 0,4 1765

TABLA X. Comparación del Muestreo Aleatorio Simple, Estratificado y Sistemático.

(Mezcla de Poblaciones: N(2,10) N(0,30), N(5,8), n=100)

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
MEDIA			
Número de Casos	100	100	100
Mínimo	1,40819	1,33762	1,57062
Máximo	3,48662	3,19938	3,13539
Mediana	2,31875	2,35497	2,30452
Media	2,36805	2,29863	2,36033
Desviación Estándar	0,41765	0,39552	0,40400
Varianza	0,17443	0,15644	0,16322
TOTAL			
Número de Casos	100	100	100
Mínimo	3.100,00	3.100,00	3.100,00
Máximo	5.900,00	5.200,00	5.100,00
Mediana	4.150,00	4.400,00	4.300,00
Media	4.234,00	4.315,00	4.227,00
Desviación Estándar	517,418	432,371	498,00
Varianza	267.000,00	187.000,00	248.000,0
PROPORCION			
Número de Casos	100	100	100
Mínimo	0,3 1000	0,1000	0,3 1000
Máximo	0,59000	0,52000	0,51000
Mediana	0,41500	0,44000	0,43000
Media	0,42340	0,43250	0,42270
Desviación Estándar	0,05174	0,04324	0,04980
Varianza	0,00268	0,00187	0,00248

Nótese que los valores a los cuales debería converger la media y la varianza de x son 2,319 y 0,20033 respectivamente, por lo que observamos que con el muestreo estratificado $\bar{x}_{est}$ se aproxima más al valor que propone el Teorema del Límite Central al igual que el $\bar{x}_{sist}$ en el muestreo sistemático. Con respecto a la estimación del total, si comparamos el valor real del total de la población que es 4.600, con cada una de las estimaciones obtenidas, podemos observar que con el muestreo estratificado obtenemos un total igual a 4.3 15 ± 432,4, con el sistemático 4.227,0 ±498,0 y con el aleatorio simple, la estimación obtenida para la proporción es igual a 4.234 ± 517,4. Como podemos observar el muestreo estratificado nos proporciona mejores estimaciones.

Para probar si la media aritmética, la media estratificada, y la sistemática se distribuyen según el teorema del límite central, realizamos la prueba K-S con la hipótesis

H_0: la media aritmética tiene una distribución N(2,3 19 , 0,20033)

VS.

H_1: -- H_0

TABLA XI. Prueba K-S para las medias x , x_{est} y $\bar{x}_{sist}$

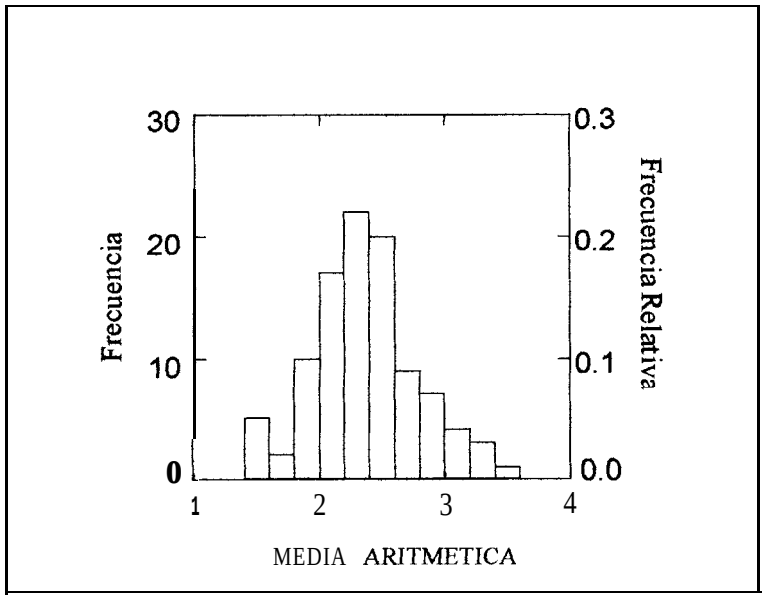
(Mezcla de Poblaciones: N(2,10) N(0,30), N(5,8), n=100)

Muestreo Aleatorio Simple			
Prueba de Kolmogorov -Smimov usando una distribución Normal(2,32 ,0,2003)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,11283	0,15673
Muestreo Estratificado			
Prueba de Kolmogorov -Smimov usando una distribución Normal(2,32, 0,2003)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,07894	0,56145
Muestreo Sistemático			
Prueba de Kolmogorov -Smimov usando una distribución Normal (2,32, 0,2003)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,09025	0,38927

Como se puede observar en la tabla XI, el valor p de la prueba para cada uno de los tipos de Muestreo nos permiten afirmar que existe evidencia estadística para aceptar la hipótesis H_0 . En la figura 3.9, 3.10 y 3.11 se presentan los histogramas de frecuencia para la media aritmética, para la media estratificada y para la media sistemática respectivamente.

FIGURA 3.9. Histograma de Frecuencia para la Media Aritmética

(Muestreo Aleatorio Simple, n=100)



Como se observa en la figura 3.9. la media aritmética está distribuida alrededor de 2,4 aproximadamente, lo cual concuerda con los resultados obtenidos en la TABLA X.

En la figura 3.10. vemos que la media sistemática ($\bar{x}_{sist}$) está distribuida alrededor de su media, esto es $2{,}3603 \pm 0{,}404$

FIGURA 3.10 Histograma de Frecuencia para la Media Sistemática.

(Muestreo Sistemático , n=100)

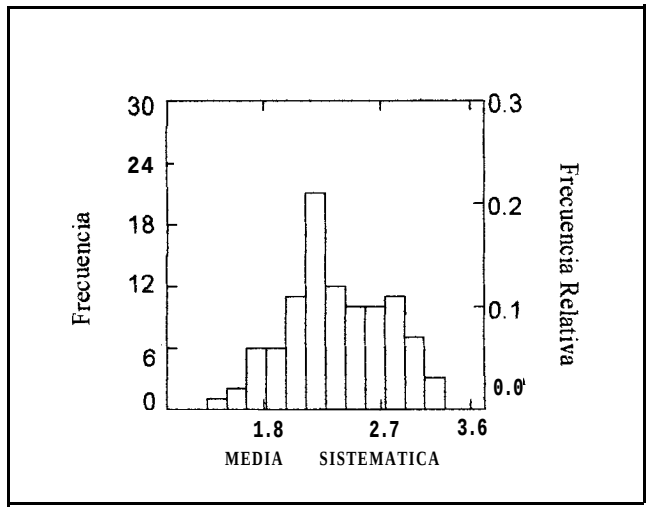
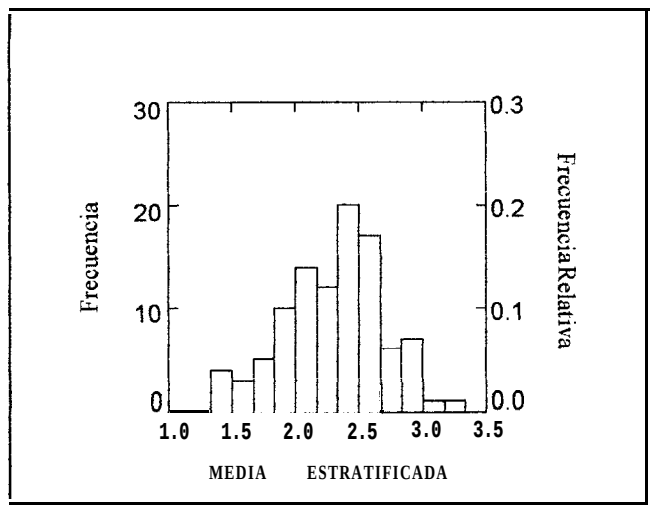


FIGURA 3.11 Histograma de Frecuencia para la Media Estratificada

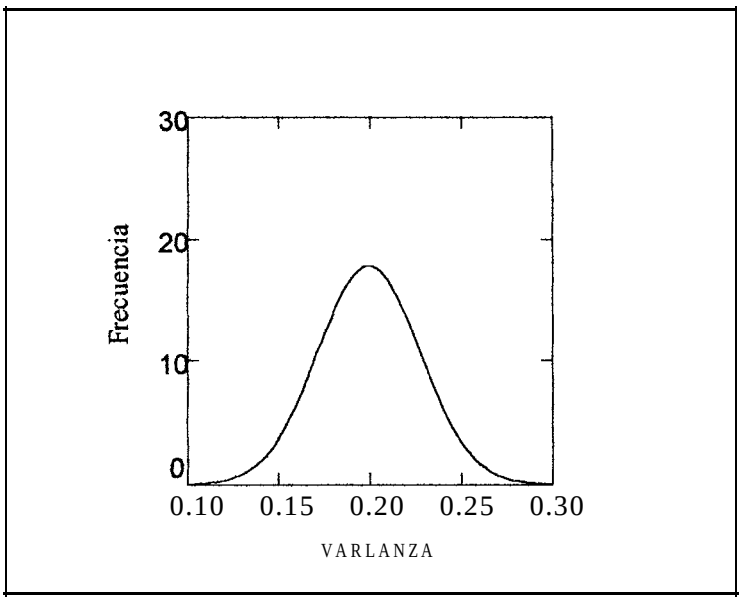
(Muestreo Estratificado, n=100)



En la figura 3.11. observamos que la media estratificada ($\bar{x}_{est}$), se distribuye alrededor de su media, es decir, de $2{,}29863 \pm 0{,}39552$.

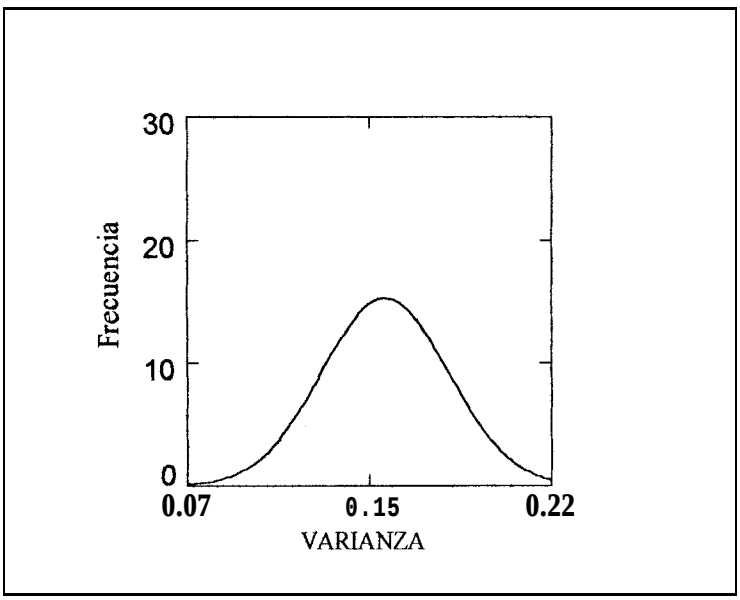
Nótese, que la dispersión que tienen los datos en cada uno de las técnicas es diferente; esto lo podemos observar de una mejor manera en los gráficos de las varianzas que a continuación se presentan.

FIGURA 3.12 Varianza de la Media Aritmética, n=100



En la figura 3.12. vemos que la **varianza** de la media aritmética se encuentra distribuida entre 0,10 y 0,30; con una media igual a 0,1744.

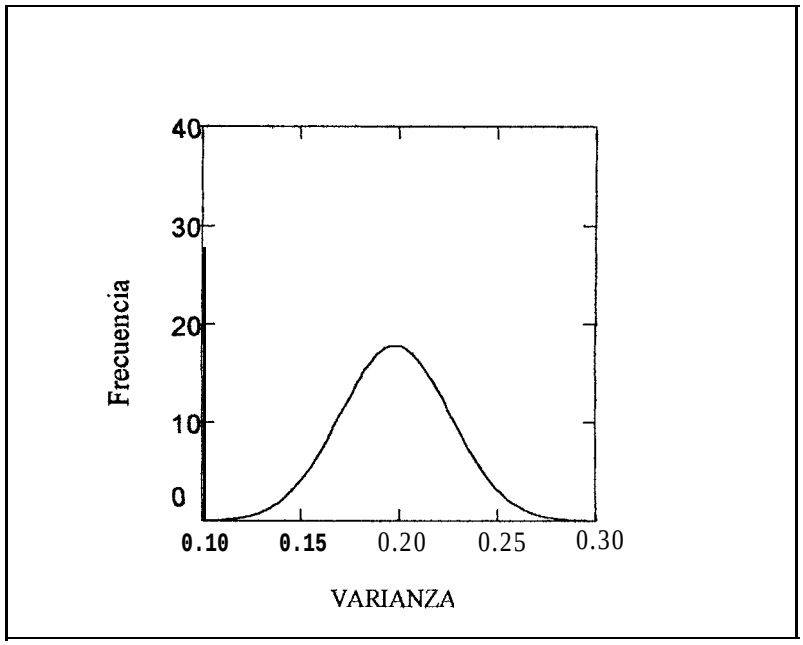
FIGURA 3.13 Varianza de la Media Estratificada, n=100



En la figura 3.13. observamos que la **varianza** de la media estratificada se encuentra distribuida entre 0,070 y 0,22, con una media igual a 0,15644.

Nótese que la **varianza** de la media estratificada es menor a la **varianza** de la media aritmética en el Muestreo aleatorio simple, como es de esperarse.

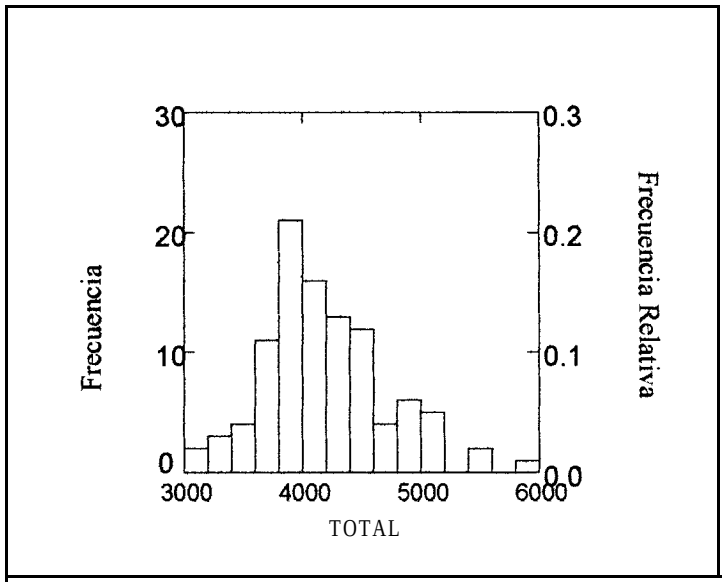
FIGURA 3.14. Varianza de la Media Sistemática, n=100



La figura 3.14. nos muestra que la **varianza** de la media sistemática se encuentra distribuida entre 0,10 y 0,25 con una media igual a 0,165322. Cabe señalar que al analizar las figuras 3.12, 3.13, y la 3.14, observamos que la técnica con la cual obtenemos una menor **varianza** para el estimador de la media, es el muestreo estratificado seguido del muestreo sistemático.

Las distribuciones que tienen el total y la proporción con el muestreo aleatorio simple, con el estratificado y con el sistemático se presentan en los siguientes gráficos.

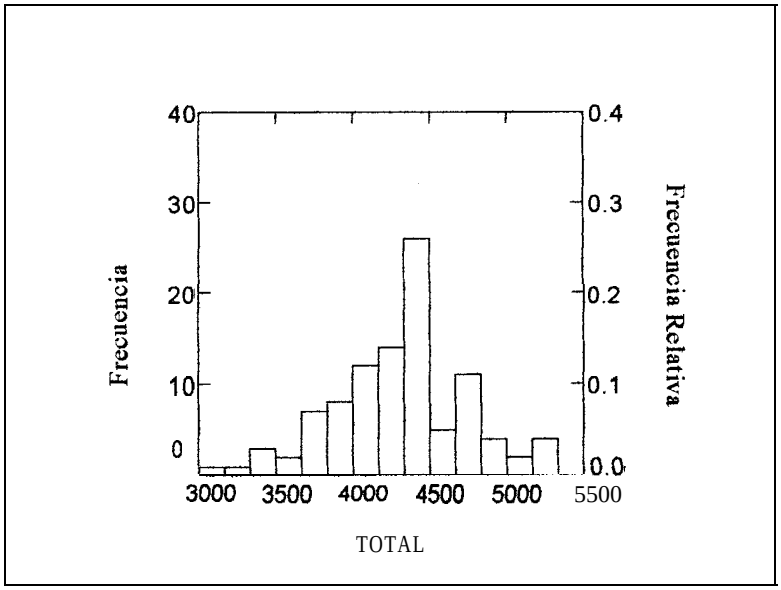
FIGURA 3.15. Histograma de Frecuencia para el Total (Muestreo Aleatorio Simple, n=100)



En la figura 3.15. podemos ver que el total en el muestreo aleatorio simple se distribuye alrededor de su media, esto es, 4.234 ± 5 17,4 . Al analizar la figura 3.16, nos damos cuenta que la distribución del total con esta técnica es diferente a la que tenía con el muestreo aleatorio simple. Este se distribuye alrededor de 4.3 15 con una desviación de 432,4.

FIGURA 3.16. Histograma de Frecuencia para el Total

(Muestreo Estratificado, n=100)

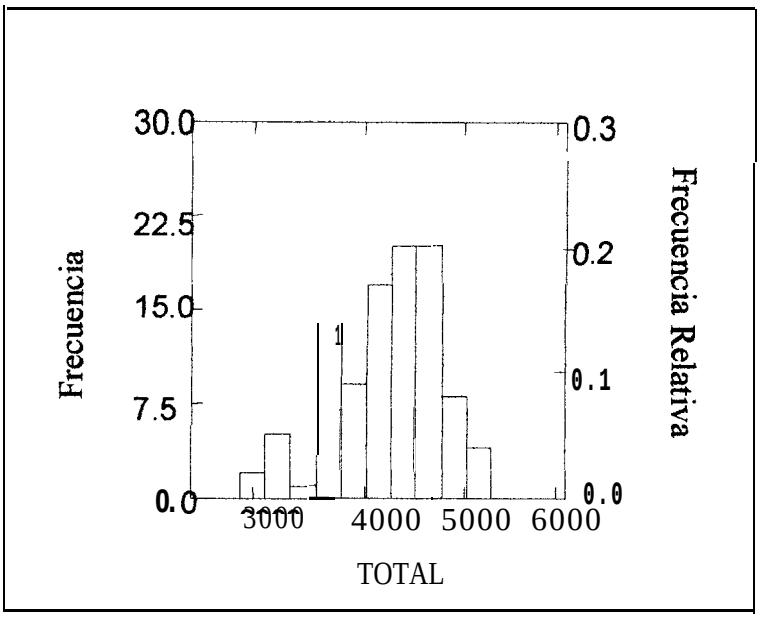


Si comparamos la figura 3.15. con la figura 3.16, podemos observar que la varianza que tiene el total en el muestreo estratificado es menor que la varianza obtenida para este estimador con el muestreo aleatorio simple.

De la figura 3.17, vemos que la distribución del total, es diferente a la que tenía con el muestreo aleatorio simple y con el estratificado. Este se distribuye alrededor de 4.227 con una desviación de 498,O.

FIGURA 3.17. Histograma de Frecuencia para el Total

(Muestreo Sistemático, n=100)



En las siguientes figuras se presentan los histogramas de frecuencia para la proporción con cada una de estas técnicas. Como podemos observar en la figura 3.18, la proporción se distribuye alrededor de 0,423 ± 0,05 174

FIGURA 3.18. Histograma de Frecuencia para la Proporción

(Muestreo Aleatorio Simple, n=100)

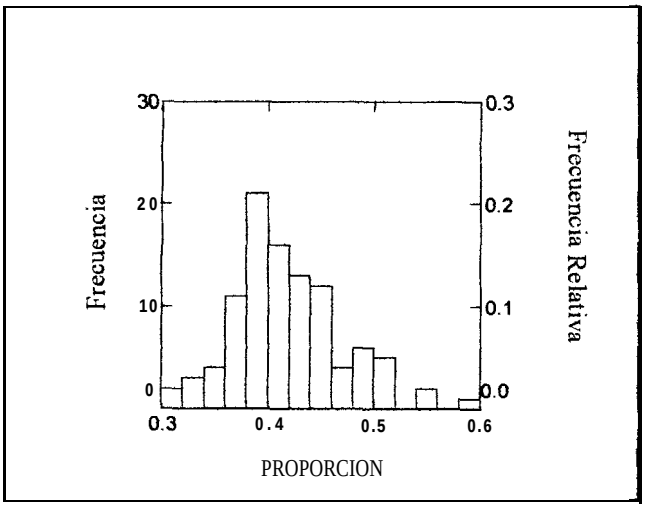
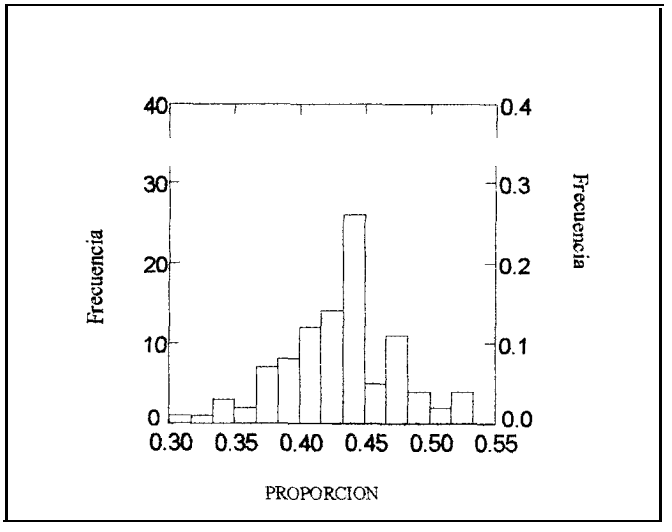


FIGURA 3.19. Histograma de Frecuencia para la Proporción

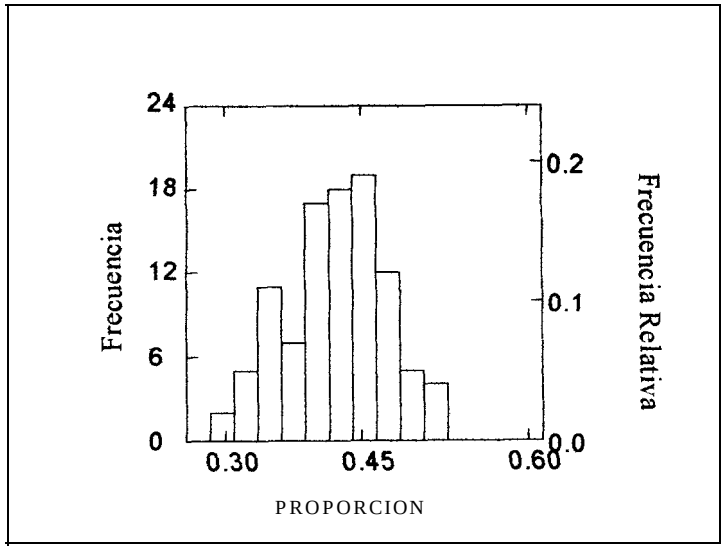
(Muestreo Estratificado, n=100)



Al analizar la figura 3.19, observamos que la proporción se distribuye alrededor de 0,43 150 con una desviación de 0,04324.

Nótese que la varianza de la proporción en el muestreo estratificado, es menor que la varianza de esta en el muestreo aleatorio simple.

FIGURA 3.20. Histograma de Frecuencia para la Proporción (Muestreo Sistemático, n=100)



Al analizar esta figura, se nota que la distribución de la proporción con esta técnica es diferente a la que tenía con el muestreo aleatorio simple y con el estratificado. Aquí la proporción se distribuye alrededor de 0,227 ± 0,0498 .

Si comparamos la figura 3.18, la figura 3.19 y la 3 .20, nos damos cuenta que la varianza que tiene la proporción en el muestreo estratificado es menor que la varianza obtenida para este estimador con el muestreo aleatorio simple y con el sistemático.

Para analizar la cobertura que tienen los intervalos de confianza con cada una de éstas tres técnicas de muestreo, construimos intervalos de confianza del 95% para la media aritmética del muestreo aleatorio simple, la media estratificada y para la media sistemática; de los cien intervalos construidos, encontramos que con el muestreo aleatorio simple, trece intervalos no contenían a la media verdadera, con el muestreo estratificado diez intervalos no contenían a la media y con el sistemático once no contenían a la media verdadera. (ver ANEXO 2).

Nótese que desde el punto de vista teórico, por lo menos noventa y cinco de los intervalos deberían contener al parámetro estimado, sin embargo al analizar las tablas del ANEXO 2, podemos observar que esto no se cumple exactamente. Así se encuentra que en el muestreo aleatorio simple ochenta y siete de los cien intervalos contienen al valor promedio verdadero, noventa en el muestreo estratificado y ochenta y nueve en el muestreo sistemático, en tanto que la teoría dice que deberían incluirlo noventa y cinco de los cien intervalos.

Al analizar los intervalos de confianza de cada una de las técnicas y calcular la longitud de sus intervalos L , definida anteriormente, encontramos que la longitud promedio de los intervalos en el muestreo aleatorio simple es de 1,745349, el estratificado es de 1,5443 y en el sistemático es de 1,7400844.

Es importante señalar que a una misma confianza, entre menos es la longitud, mayor información nos proporciona el intervalo. Al analizar esto podemos concluir que con muestreo estratificado se obtiene una mejor estimación para la media, que con el muestreo aleatorio simple y sistemático.

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

CAPITULO IV

COMPARACIÓN DE LOS RESULTADOS OBTENIDOS CON LOS DIFERENTES TIPO DE MUESTREO UTILIZANDO POBLACIONES DISCRETAS Y CONTINUAS.

4.1. Generalidades

En el presente capítulo analizaremos el comportamiento que toman algunos estimadores como la media y el total, para casos en que la población es discreta y también cuando ésta es continua. Para el efecto trabajaremos con poblaciones que siguen una distribución de Poisson para el caso discreto y Normal para el caso continuo. Cabe señalar que este estudio estará enfocado en analizar las distribuciones de probabilidad de la media y del total para cada uno de los casos, así como también analizaremos las varianzas de los estimadores, el error de estimación obtenido con cada uno de las técnicas de muestreo y un intervalo de confianza para dichos estimadores.

Antes de continuar con el desarrollo de este capítulo, es necesario recordar las definiciones de las dos distribuciones de probabilidad a utilizarse.

Distribución de Poisson.

Una variable aleatoria X tiene una distribución de Poisson y se denomina variable aleatoria de Poisson con parámetro λ , si y sólo si su distribución de probabilidades está dada por

$$p(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{para } x = 0, 1, 2, 3, \ldots..$$

$\lambda > 0$

con media

$$\mu = \lambda$$

y varianza $\sigma^2 = \mu = \lambda$

como probaremos a continuación

$$E(x) = \sum_{x=0}^{\infty} \frac{x e^{-\lambda} \lambda^x}{x!}$$

$$= \sum_{x=1}^{\infty} \frac{x e^{-\lambda} \lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{x \lambda^{-}}{(x-1)!}$$

$$= \lambda e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\ldots\ldots\ldots\right)$$

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda$$

De igual manera se puede demostrar que

$$\sigma^2 = E\left(x^2\right) - \mu^2$$

$$= \mathrm{E}\left(\mathrm{x}^2\right) - \lambda^2 = \lambda$$

Distribución Normal

Una variable aleatoria X tiene una distribución de normal y se conoce como variable aleatoria normal, si y sólo si su distribución de probabilidad está dada por

$$\mathrm{f(x) = n\ (x;\mu;\sigma)} = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\mathrm{x}-\mu}{\sigma}\right)^2} \quad \text{para } -\infty < x < \infty$$

donde $\sigma > 0$ y $-\infty < \mu < \infty$

con media μ y varianza σ^2 .

Una vez que hemos revisado estas definiciones, pasaremos a realizar el estudio para cada una de las técnicas de muestreo. En la primera parte de este capítulo compararemos los resultados obtenidos con el muestreo aleatorio simple, el estratificado y el sistemático utilizando una población Poisson, y en la segunda parte, realizaremos un análisis de éstas tres técnicas, pero cuando la población es Normal.

4.2 Análisis de los Resultados obtenidos con el Muestreo Aleatorio Simple, con el Estratificado y con el Muestreo Sistemático utilizando una población Discreta

Para realizar éste análisis se generaron 3.333 valores de una variable aleatoria con distribución de Poisson con $\lambda = 2$; 3.333 valores de una variable aleatoria con distribución de Poisson con $\lambda = 5$ y 3.334 de una Poisson con $\lambda = 10$ Estos valores pasaron a conformar el marco muestral. Nótese que vamos a muestrear de una población que contiene a su vez subpoblaciones cada una con diferente media y varianza. Definido el marco muestral, calculamos la media y la varianza de esta población, de lo cual obtuvimos

$$\mu = 5{,}6566 \quad \text{y } \sigma^2 = 16{,}67647$$

Además de la media y de la varianza, calculamos el total de elementos en la población que eran menores o iguales 5, obteniéndose

$$A = 5.555$$

Una vez construido el marco, con cada una de las técnicas tomamos 100 muestras cada una de ellas de tamaño 100; en el muestreo estratificado para determinar el tamaño de la muestra para cada estrato se utilizó la asignación proporcional.

Luego de haber obtenido las muestras, realizamos las estimaciones antes mencionadas, es decir, estimamos el total, la media y calculamos la varianza de cada uno de éstos estimadores. Para el análisis de la distribución de X utilizamos la prueba de Kolmogorov – Smirnov (K- S) la misma que como se mencionó en el capítulo anterior nos ayudará a determinar la distribución que sigue la media. A continuación se presentan los resultados obtenidos con los diferentes tipos de muestreo.

TABLA XII. Cuadro Comparativo del estimador de la media poblacional,

(Mezcla de Poblaciones: P(2), P(5), P(10), n=100)

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
MEDIAS			
Número de Casos	100	100	100
Mínimo	5,04000	5,03000	5,09000
Máximo	7,35000	6,26000	7,55000
Mediana	6,17500	5,70000	6,19500
Media	6,18970	5,69550	6,16530
Desviación Estándar	0,48496	0,24256	0,42873
Varianza	0,23518	0,05884	0,18381

Como podemos observar el la TABLA XII, el muestreo estratificado nos proporciona una mejor estimación para la media poblacional, ya que los otros dos métodos sobrestiman éste parámetro. Con el muestreo estratificado observamos que la media estimada es de 5,695 ± 0,24256, mientras que en el muestreo aleatorio simple y en el sistemático es de 6,1897 ± 0,48496 y 6,1653 ± 042873 respectivamente.

Nótese que

$$\frac{\sigma}{\sqrt{n}} = 0{,}406 = \sigma_{\bar{x}}$$

Siendo μ = 5,6566 , podemos observar que el muestreo aleatorio simple y el sistemático la sobrestiman, mientras que el muestre estratificado la subestima. Así como hemos analizado la varianza que tiene este estimador, es necesario construir un intervalo de confianza del 95% para la media, ya que de esta manera podemos establecer con cuál técnica se obtiene un intervalo que proporcione mayor información, con igual confianza.

Intervalo para la Media del Muestreo Aleatorio Simple

Límite Superior 6,28593

Límite Inferior 6,09347

Intervalo para Media del Muestreo Estratificado

Limite Superior 5,74363

Límite Inferior 5,64737

Intervalo para la Media del Muestreo Sistemático.

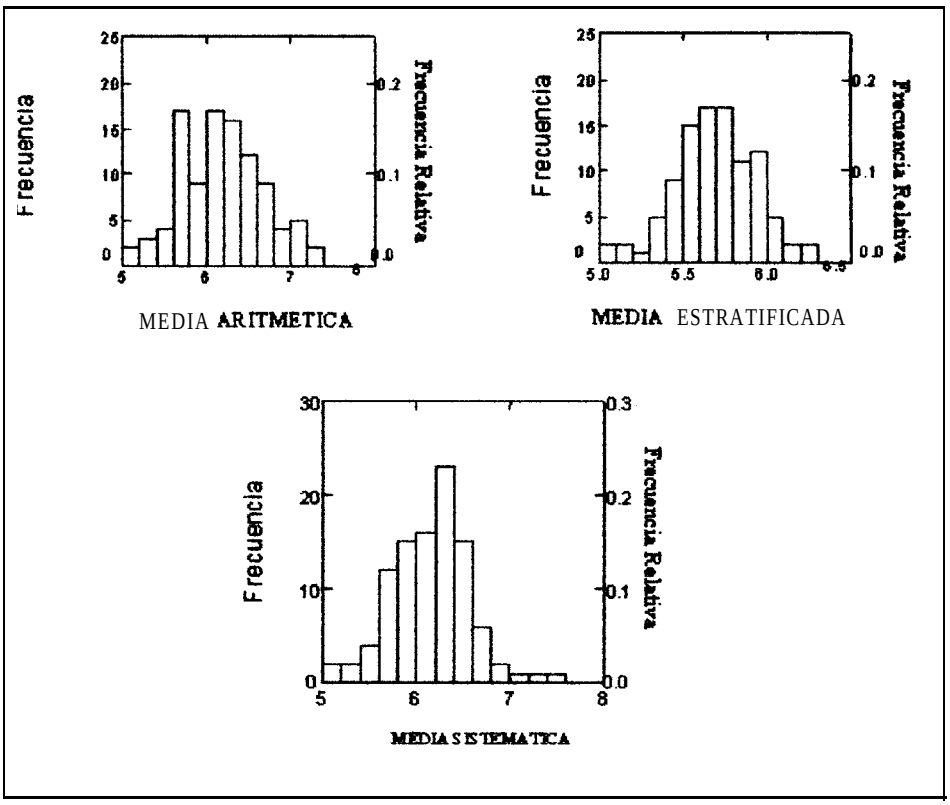
Límite Superior 6,25037

Límite Inferior 6,08023

Como podemos observar con el muestreo estratificado, obtenemos un interv de confianza que nos proporciona mayor información, ya que su longitud menor que la de los otro dos.

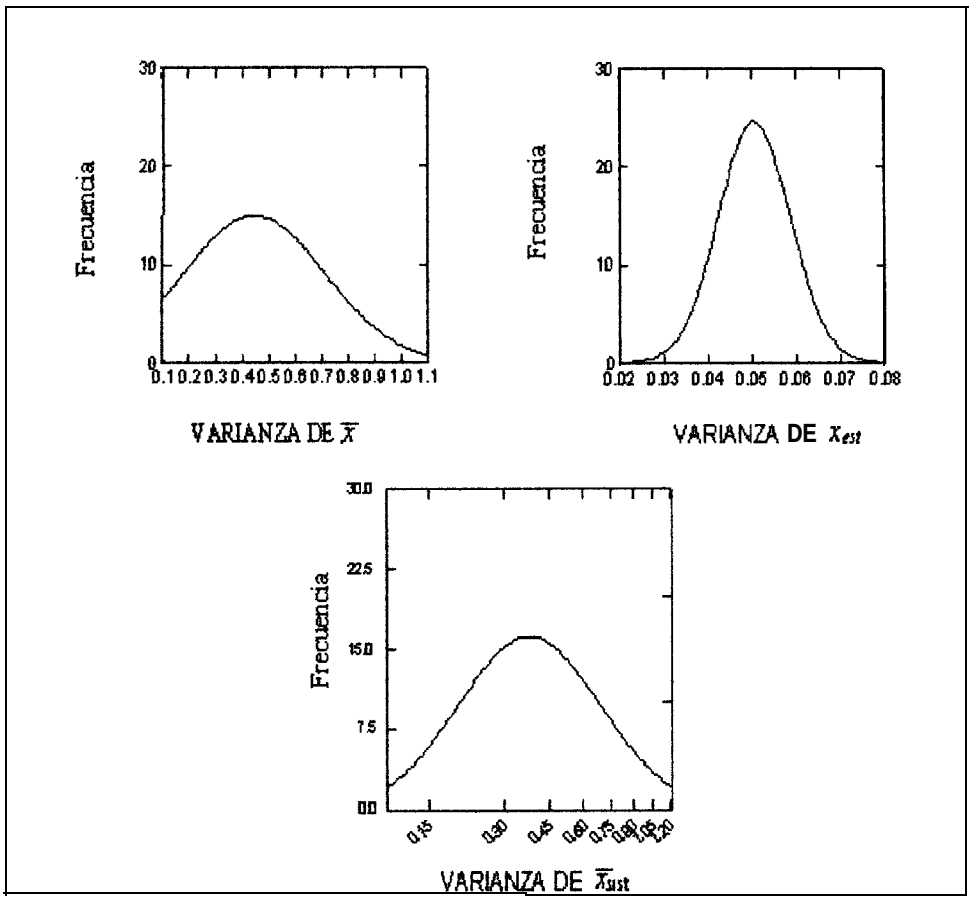
En los gráficos siguientes se presentan los histogramas de **frecuencia** para media aritmética del muestreo aleatorio simple, para la media estratificada y pa la media sistemática respectivamente.

FIGURA 4.1 Histograma de Frecuencia para la Media Aritmética, para la Media Estratificada y para la Media Sistemática, n=100



Como podemos observar en la figura 4.1, la media estratificada se encuentra distribuida en una forma más simétrica que las otras dos, se puede observar también en cada uno de los histogramas que las varianzas de los estimadores no son iguales aunque los tres se encuentran distribuidos al rededor de 5. Esto se presenta en una forma más clara en los gráficos de las varianzas de estos estimadores.(figura 4.2).

FIGURA 4.2. Varianza de los Estimadores



Como podemos observar en esta figura 4.2, con el muestreo estratificado, obtenemos una menor **varianza** que con las otras técnicas de muestreo, esto se puede ver en la amplitud de la "campana" en cada una de los casos. Para verificar si la media aritmética, la media estratificada, y la sistemática tiene una distribución normal, realizamos la prueba K-S con las hipótesis

H_0: tiene una distribución N(5,66 , 0,166764)

VS.

H_1: $\neg H_0$.

TABLA XIII. Prueba K - S para las medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$

(Mezcla de Poblaciones: P(Z), P(5), P(10), n=100)

Muestreo Aleatorio Simple			
Prueba de Kolmogorov - Smimov **usando una** distribución **Normal(5.66,0. 1667)**			
Variable	N de Casos	Máxima Diferencia	Valor p **(2-colas)**
MEDIA	100	0,46860	0,00210
Muestreo Estratificado			
Prueba de Kolmogorov -Smimov usando una distribución **Normal(5.66,0. 1667)**			
Variable	N de Casos	Máxima Diferencia	Valor p **(2-colas)**
MEDIA	100	0,17645	0,00395

Muestreo Sistemático			
Prueba de Kolmogorov -Smimov usando una distribución Normal (5.66,0. 1667)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,46255	0,00259

Como se puede observar en la TABLA XIII, el valor p de la prueba en los tres tipos de muestreo es pequeño, por lo que en este caso no se puede aceptar la hipótesis nula de que las medias tienen una distribución N (5,6566, 0,166764).

Nótese entonces que en el caso de poblaciones finitas, no siempre con un tamaño de muestra mayor a 30 y esta cantidad de muestras (100), la media se va a distribuir con media μ y varianza $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

A continuación presentamos los resultados obtenidos cuando estimamos el total poblacional con cada una de las técnicas de muestreo.

TABLA XIV. Cuadro Comparativo del estimador del total poblacional.

(Mezcla de Poblaciones: P(2), P(5), P(10), n=100)

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
TOTAL			
Número de Casos	100	100	100
Mínimo	4300,00000	4646,00000	4800,00000
Máximo	7000,00000	6443,41176	6600,00000
Mediana	5650,00000	5547,64706	5700,00000
Media	5584,00000	5525,98588	5613,00000
Desviación Estándar	489,22119	332,99189	377,03 120
Varianza	2,39337E+05	1,10884E+05	1,42153E+05

Al analizar esta TABLA XIV y recordar que el total poblacional es igual a 5.555, podemos observar que la técnica que nos da un estimador con menor varianza es el muestreo estratificado como era de esperarse, aunque no es el que mejor estima al parámetro. Para poder determinar de una mejor manera cual técnica nos proporciona la mejor estimación calculamos el error obtenido en cada una de las técnicas, a continuación se presentan los resultados obtenidos.

ESP
ESCUELA SUPERIOR POLITECNICA DEL LITO
CIB - ESPO

Muestreo Aleatorio Simple

$e = |\hat{\theta} - \theta|$

$e = |5584 - 5555|$

$e = 29$

Muestreo Estratificado

$e = 15525 - 5555|$

$e = 30$

Muestreo Sistemático

$e = |5613 - 5555|$

$e = 58$

Al analizar estos resultados podemos observar que el muestreo estratificado aunque proporciona un estimador de menor **varianza,** el error de estimación que se obtiene es mayor que el obtenido con el muestreo aleatorio simple.

Antes de continuar con la segunda parte de éste capítulo, analizaremos el comportamiento del estimador de la media cuando la población de la que se muestrea está constituida por un solo tipo de variable aleatoria. En este estudio

realizaremos estimaciones de la media poblacional con el muestreo ales simple y con el sistemático partir de una población con distribución Poisso λ = 2, 5 y 10 ; para cada uno de los casos la población generada fue de tai 10.000 . La tabla que se presenta a continuación muestra la estadí, descriptiva para el estimador de la media poblacional con cada una de técnicas.

De esta tabla podemos observar que tanto el muestreo sistemático como aleatorio simple ofrecen 'buenas" estimaciones para la media poblacional e cada uno de los casos ya que no hay diferencia mayor entre las dos estimaciones es así como, para la población con λ = 2 se obtiene un valor de 2,0049 ± 0,12965 con el muestreo sistemático y con el aleatorio simple el estimador es igual a 1,9865 ± 0,13 14 .

De la misma manera en la población con λ = 5 con el muestreo sistemático obtenemos un valor de 5,0132 ± 0,2245 para el estimador de la media, mientras que con el aleatorio simple el estimador de la media poblacional es igual a 4,984 ± 0,22282 . Por último en la población con λ = 10, el estimador de la media con el muestreo sistemático es igual a 10,05 16 ± 0,3 1937 y en el muestreo aleatorio simple es de 9,975 ± 0,3 1447 .

TABLA XV. Comparación de los resultados con Muestreo Aleatorio Simple y Sistemático, n=100.

Población Poisson con $\lambda = 2$	**Muestreo Aleatoria Simple**	**Muestreo Sistemático**
MEDIA		
Número de Muestras	100	100
Mínimo	1,60000	1,64000
Máximo	2,32000	2,28000
Mediana	1,97500	1,99500
Media	1,98650	2,00490
Desviación Estándar	0,13144	0,12965
Varianza	0,01728	0,01681
Población Poisson con $\lambda = 5$	**Muestreo Aleatorio Simple**	**Muestreo Sistemático**
MEDIA		
Número de Muestras	100	**100**
Mínimo	4,40000	4,33000
Máximo	5,73000	5,47000
Mediana	4,97000	5,00500
Media	4,98400	5,01320
Desviación Estándar	0,22288	0,22458
Varianza	0,04967	0,05044
Población Poisson con $\lambda = 10$	**Muestreo Aleatorio Simple**	**Muestreo Sistemático**
MEDIA		
Número de Muestras	**100**	**100**
Mínimo	9,11000	9,12000
Máximo	10,69000	10,88000
Mediana	9,96000	10,04500
Media	9,97500	10,05160
Desviación Estándar	0,3 1447	0,3 1937
Varianza	0,09889	0,10200

Para verificar si la media tenía una distribución N(2, 0,20) en la población con $\lambda = 2$, N(5, 0,50) en la población con $\lambda = 5$ y N(10, 0,10) en la población con $\lambda = 10$, se establecieron las siguientes pruebas de hipótesis.

Para el primer caso:

H_0: la media aritmética tiene una distribución N(2 , 0,.20)

VS.

H_1: $\neg H_0$.

Para el segundo caso:

H_0: la media aritmética tiene una distribución N(5 , 0,50)

VS.

H_1: $\neg H_0$.

Para el tercer caso:

H_0: la media aritmética tiene una distribución N(10 , 0,10)

VS.

H_1: $\neg H_0$.

TABLA XVI. Prueba K-S para las medias $\bar{x}$ y $\bar{x}_{sist}$

(Poblaciones Discretas)

Población con λ = 2

Prueba de Kolmogorov – Smimov usando una distribución **Normal(2,0 ,0,2)**

Media del Muestreo Aleatorio Simple

N de Casos	Máxima Diferencia	Valor p (2-colas)
100	0,11000	0,17772

Media del Muestreo Sistemático

N de Casos	Máxima Diferencia	Valor p (2-colas)
100	0,11482	0,14312

Población con λ = 5

Prueba de Kolmogorov -Smimov usando una distribución **Normal(5., 0,50)**

Media del Muestreo Aleatorio Simple

N de Casos	Máxima Diferencia	Valor p (L-colas)
100	0,10212	0,24801

Media del Muestreo Sistemático.

N de Casos	Máxima Diferencia	Valor p (2-colas)
100	0,09792	0,28742

Población con λ = 10

Prueba de Kolmogorov – Smimov usando una distribución Normal (10.0, 0,10)

Media del Muestreo Aleatorio Simple

N de Casos	Máxima Diferencia	Valor p (2-colas)
100	0,09779	0,29444

Media del Muestreo Sistemático.

N de Casos	Máxima Diferencia	Valor p (2-colas)
100	0,08950	0,39963

Como se puede observar en la TABLA XVI, el valor p de la prueba en cada uno de los casos permite aceptar todas las hipótesis nulas propuestas anteriormente. Es importante señalar que si realizamos una comparación entre estos resultados y los obtenidos cuando la población estaba constituida por tres subpoblaciones podemos observar que la convergencia de la media es mucho más rápida cuando la población es "pura" que cuando ésta es el resultado de una mezcla de varias poblaciones.

4.3 Análisis de los Resultados obtenidos con el Muestreo Aleatorio Simple, con el Estratificado y con el Muestreo Sistemático utilizando una población Continua

En esta parte del estudio analizamos el comportamiento de los estimadores antes mencionados, cuando la población de la cual se muestrea, está constituida por tres subpoblaciones de igual media pero diferente varianza, y una población con tres subpoblaciones que tienen igual varianza pero diferente media.

Para realizar el análisis en la primera población, se generaron 3.333 valores de una variable aleatoria Normal (2, 7); 3.333 valores de una Normal(2, 24); y 3.334 de una Normal(2,13). Estos valores pasaron a formar el marco muestral.

Definido el marco muestral, calculamos la media y la varianza de esta población, de lo cual obtuvimos

$$\mu = 2{,}0092 \quad \text{y} \quad \sigma^2 = 14{,}6641$$

Además de la media y de la varianza, calculamos el total de elementos en la población que eran menores o iguales 5, obteniéndose

$$A = 4.981$$

Una vez construido el marco, con cada una de las técnicas tomamos 100 muestras cada una de ellas de tamaño 100; en el muestreo estratificado al igual que en los casos anteriores, para determinar el tamaño de la muestra para cada estrato realizamos una asignación proporcional.

Luego de haber obtenido las muestras, realizamos las estimaciones antes mencionadas. Para el análisis de la distribución de X utilizamos la prueba K- S.

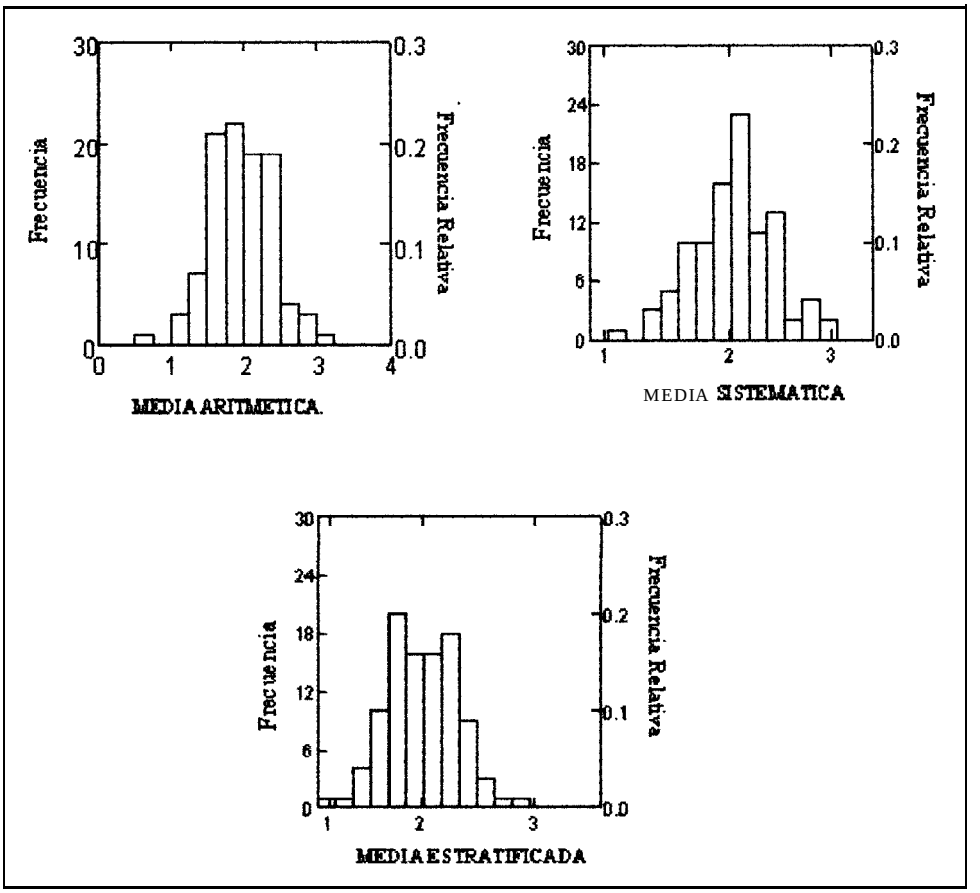
A continuación se presentan los resultados obtenidos con los diferentes tipos de muestreo.

TABLA XVII. Comparación de los resultados con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático. (Mezcla de Poblaciones: N(2,7) , N(2,24) , N(2,13) ,n=100)

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
MEDIA			
Número de Casos	100	100	100
Mínimo	0,8501	1,06421	1,12814
Máximo	3,09571	2,93875	2,87449
Mediana	1,94587	2,00361	2,04201
Media	1,96880	2,00763	2,04648
Desviación Estándar	0,4 1006	0,33601	0,36621
Varianza	0,16815	0,11290	0,13411

En la TABLA XVII podemos observar que el muestreo estratificado al igual que el sistemático ofrecen una mejor estimación para la media poblacional que el muestreo aleatorio simple, es así que con el estratificado se obtiene un valor de 2,00763 ± 0,336 y con el muestreo sistemático una estimación de 2,04648 ± 0,36621, mientras que con el aleatorio simple el estimador es igual a 1,96880. La distribución que tiene el estimador de la media poblacional en cada uno de las técnicas se puede observar de una mejor manera en los histogramas de frecuencia que a continuación se presentan.

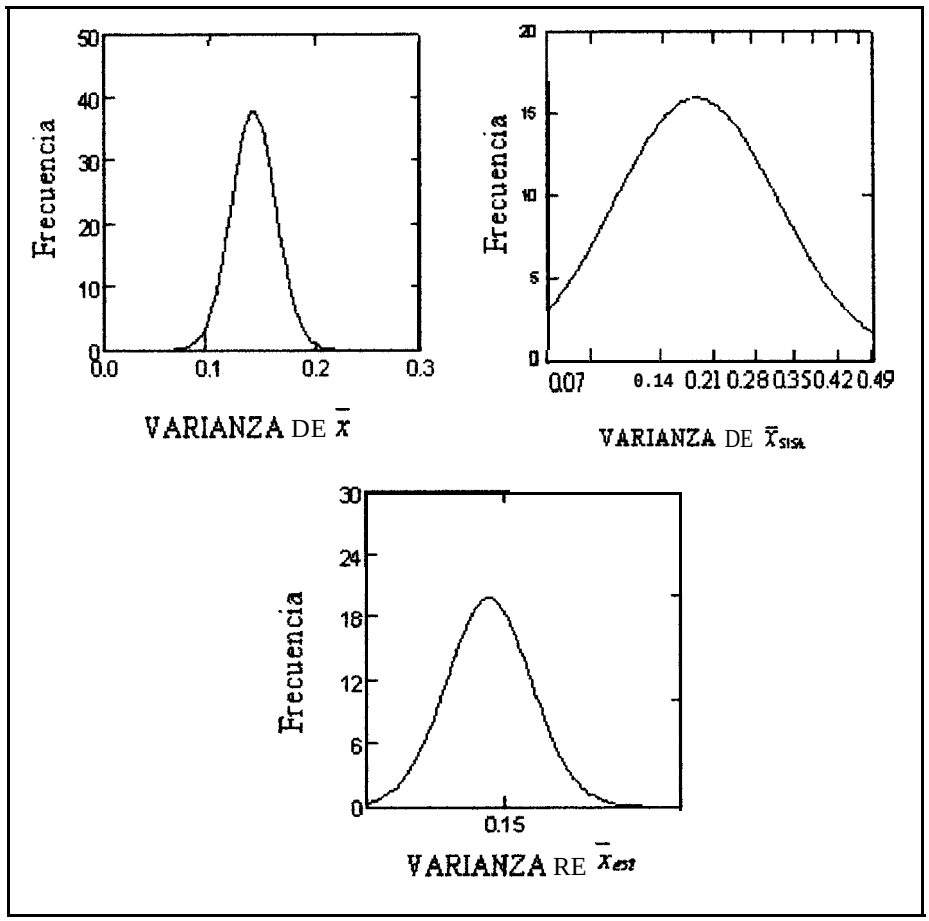
FIGURA 4.3 Histograma de Frecuencia para x, $\bar{x}_{sist}, \bar{x}_{est}$



Si miramos en la figura 4.3 la distribución que tiene la media en cada una de las técnicas, nos damos cuenta que la media estratificada tiene una menor dispersión que las otras dos, aunque todas se concentran alrededor de la misma media, es decir, 2.

La dispersión que tiene cada uno de estos estimadores se la puede observar en la figura 4.4

FIGURA 4.4 Varianza de los Estimadores de la Media



Podemos observar que la varianza que tiene el estimador de la media poblacional en el muestreo estratificado es menor que en las otras dos técnicas.

Un intervalo de confianza del 95% para la media de cada uno de los métodos estará dado por,

h 1 . 9 6 ;

De lo cual se obtiene un límite superior y uno inferior, a continuación se presentan los intervalos obtenidos.

Intervalo para la Media del Muestreo Aleatorio Simple

Límite Superior	2,05017
Límite Inferior	1,88744
Desviación Estándar	0,41006

Intervalo para la Media del Muestreo Estratificado

Límite Superior	2,07430
Límite Inferior	1,94096
Desviación Estándar	0,33601

Intervalo para la Media del Muestreo Sistemático.

Límite Superior	2,11915
Límite Inferior	1,97382
Desviación Estándar	0,36621

Al analizar cada uno de estos intervalos notamos que el que tiene menor longitud es el obtenido con el muestreo estratificado, es decir que este intervalo nos provee mayor información que el obtenido con el muestreo aleatorio simple y el sistemático.

Para verificar si la media aritmética del muestreo aleatorio simple, la media estratificada, y la sistemática tiene una distribución normal, realizamos la prueba K-S con las siguientes hipótesis

H_0: la media aritmética tiene una distribución $N(2{,}009\ ,\ 14{,}664\ 1)$

VS.

H_1: $\neg\ H_0$.

TABLA XVIII. Prueba K-S para las medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$

(Mezcla de Poblaciones: N(2,7) , N(2,24) , N(2,13))

Muestreo Aleatorio Simple			
Prueba de Kolmogorov – Smimov usando una distribución Normal (2,0, 14,66)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,09052	0,3 7936
Muestreo Estratificado			
Prueba de Kolmogorov -Smimov usando una distribución Normal (2,0, 14,.66)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,07267	0,66651
Muestreo Sistemático			
Prueba de Kolmogorov – Smimov usando una distribución Normal (2,0, 14,66)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,09043	0,38675

Como se puede observar en la TABLA XVIII, el valor p de la prueba en los tres tipos de muestreo nos permite afirmar que existe evidencia estadística para aceptar la hipótesis nula propuesta anteriormente. A continuación presentamos los resultados obtenidos cuando estimamos el total poblacional con cada una de las técnicas de muestreo.

TABLA XIX. Comparación de los resultados con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático. (Mezcla de Poblaciones: N(2,7) , N(2,24) , N(2,13), n=100)

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
TOTAL			
Número de Casos	100	100	100
Mínimo	3800,00000	4105,70588	4000,0000
Máximo	6100,00000	6295,3 5294	6200,0000
Mediana	4900,00000	4910,76471	5100,0000
Media	4932,00000	4945,00882	5023,0000
Desviación Estándar	492,95645	439,97916	465,98543
Varianza	2,43006E+05	1,93582E+05	2,17142E+05

Si recordamos que el total poblacional es 4.981 y analizamos esta tabla podemos observar que con el muestreo estratificado y el muestreo aleatorio simple subestiman al parámetro, mientras que el muestreo sistemático lo sobrestima; con el muestreo estratificado el estimador toma un valor de 4.945 ± 439,97 , en el sistemático un valor de 5.023 ± 465,98 y con el muestreo aleatorio simple el estimador es igual a 4.932 ± 492,95 .

Para determinar cual técnica nos proporcionaba un menor error en la estimación realizamos los siguientes cálculos.

Muestreo Aleatorio Simple

$$e = |\hat{\theta} - \theta|$$

$$e = |4932 - 4981|$$

$$e = 49$$

Muestreo Estratificado

$$e = |4945 - 4981|$$

$$e = \mathbf{36}$$

Muestreo Sistemático

$$e = |5023 - 4981|$$

$$e = 42$$

Al analizar estos resultados podemos observar que el estimador obtenido con el muestreo estratificado da una mejor estimación de la media poblacional que el aleatorio simple y sistemático, esto se puede observar en el error de estimación obtenido para cada caso. En las figuras 4.5, 4.6 y 4.7 se muestran los histogramas de frecuencia para el total obtenido con muestreo aleatorio simple, del muestreo estratificado y del sistemático respectivamente

FIGURA 4.5 Histograma de Frecuencia para el Total

(Muestreo Aleatorio Simple, n=100)

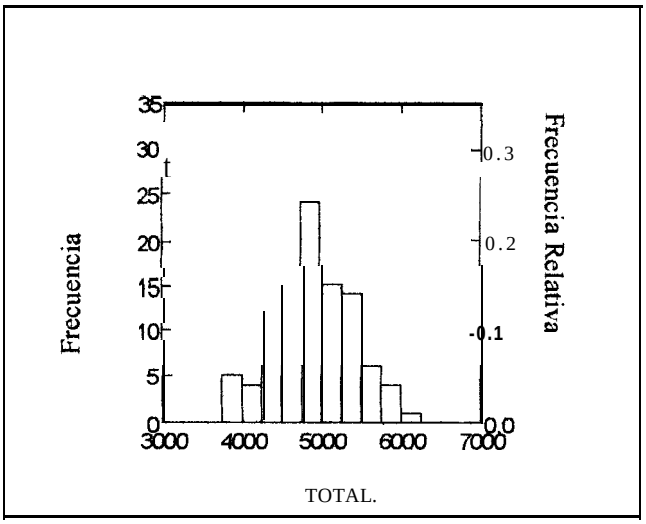


FIGURA 4.6 Histograma de Frecuencia para el Total

(Muestreo Sistemático, n=100)

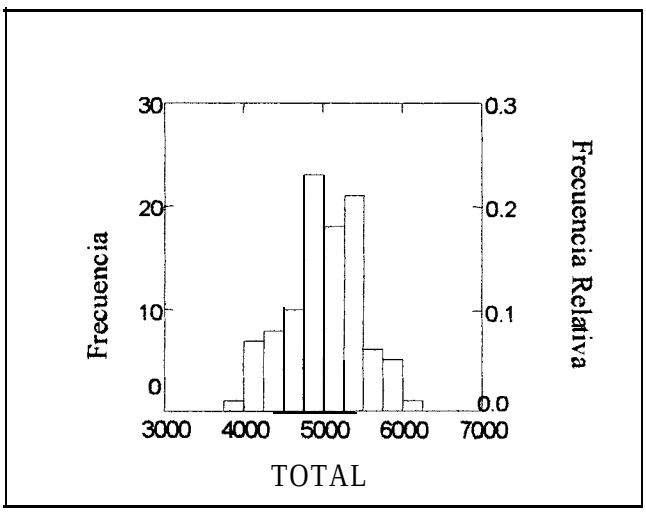
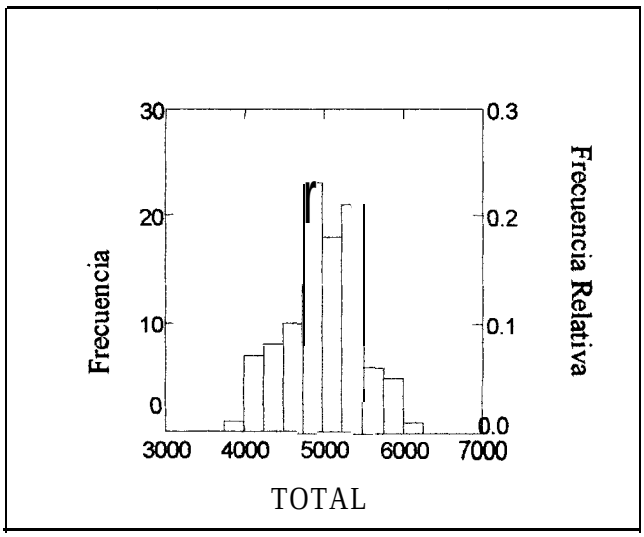


FIGURA 4.7 Histograma de Frecuencia para el Total

(Muestreo Estratificado, n=100)



En cada una de las figuras presentadas se puede observar que la dispersión del total es diferente para cada uno de los casos; en el muestreo estratificado el estimador del total tiene una menor dispersión que en los otros dos métodos.

A continuación se presenta un análisis similar al anterior, pero considerando una población constituida a su vez por tres subpoblaciones con medias diferentes, pero con la misma varianza. Para dicho efecto se generaron 3.333 valores de una variable aleatoria Normal (4, 15); 3.333 valores de una Normal(2, 15); y 3.334 de una Normal(8, 15).

Con estos valores se formó un marco muestral, del cual más adelante se procederá a seleccionar muestras. Definido el marco muestral, calculamos la media y la varianza de esta población, así como también el total de elementos en la población que eran menores o iguales 4.66, obteniéndose

$$\mu = 4{,}6677 \quad , \quad \sigma^2 = 21{,}1182 \text{ y } A = 5.019$$

Una vez construido el marco, con cada una de las técnicas tomamos 100 muestras cada una de ellas de tamaño 100; al igual que en los casos anteriores para determinar el tamaño de la muestra para cada estrato realizamos una asignación proporcional en el muestreo estratificado.

Luego de haber obtenido las muestras, realizamos las estimaciones antes mencionadas así como también la prueba K- S. A continuación se presentan los resultados obtenidos con los diferentes tipos de muestreo.

En la TABLA XX podemos observar que el estimador del muestreo sistemático al igual que el del aleatorio simple ofrece una estimación más cercana a μ que el estratificado, sin embargo si analizamos las varianzas de cada uno de estos estimadores, podemos observar que en el muestreo estratificado el estimador tiene una menor varianza que en los otros casos; así tenemos que con el muestreo

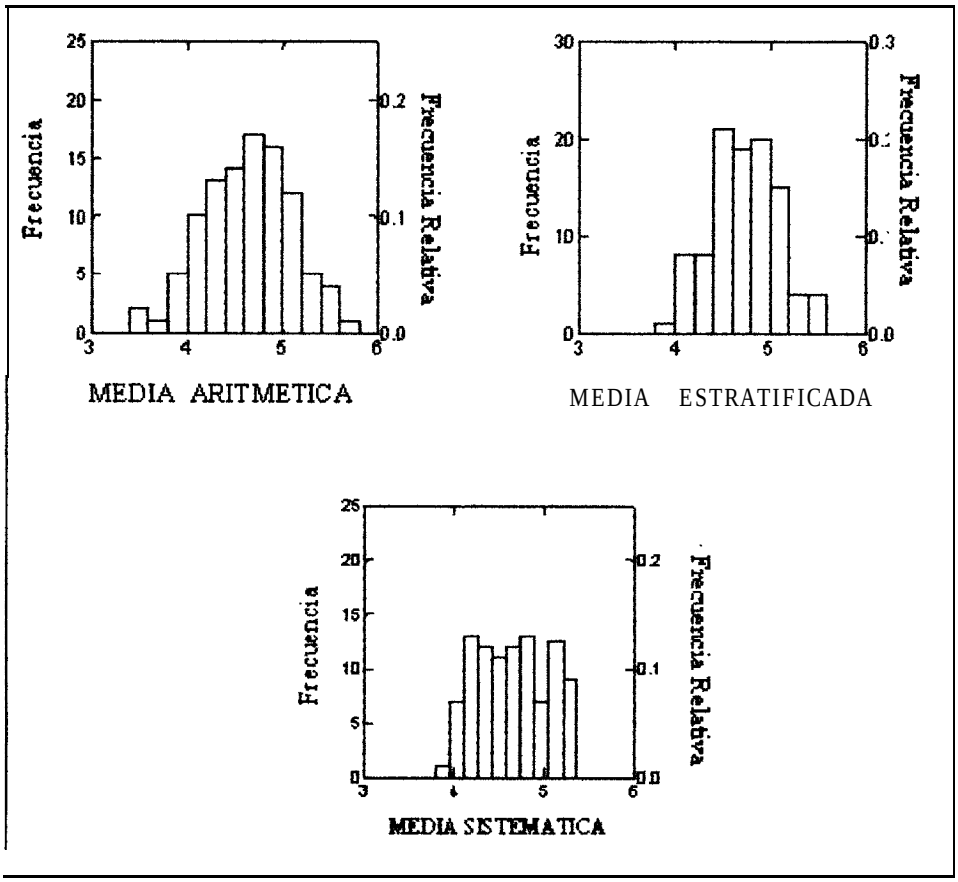
aleatorio simple el estimador toma el valor de $4{,}63418 \pm 0{,}45871$, con el estratificado, es igual a $4{,}72461 \pm 0{,}36289$ y con el sistemático es igual a $4{,}66068 \pm 0{,}38488$.

TABLA XX. Comparación de los resultados con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático. (Mezcla de Poblaciones: N(4,15), N(2,15) , N(8,15), n=100)

	Muestreo Aleatorio Simple	**Muestreo Estratificado**	**Muestreo Sistemático**
MEDIA			
Número de Casos	**100**	100	100
Mínimo	3,49738	3,91057	3,92493
Máximo	5,74030	5,58279	5,28830
Mediana	4,62457	4,74863	4,66651
Media	4,63418	4,72461	4,66068
Desviación Estándar	0,45871	0,36289	0,38488
Varianza	0,21041	0,13169	0,14813

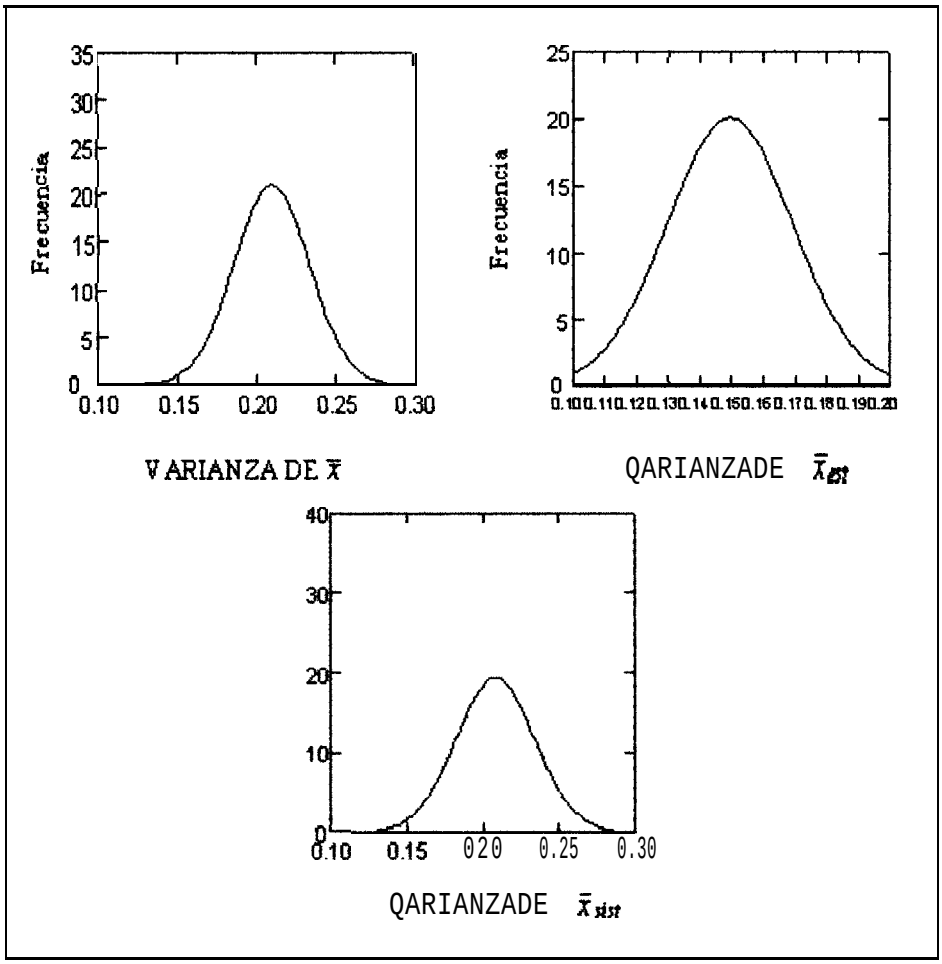
La distribución que tiene el estimador de la media poblacional en cada uno de las técnicas se puede observar de una mejor manera en los histogramas de frecuencia que a continuación se presentan.

FIGURA 4.8 Histograma de Frecuencia para x, x_{sist}, $\bar{x}_{est}$



Si miramos la distribución que tiene la media en cada una de las técnicas, nos damos cuenta que la media estratificada tiene una menor dispersión que las otras dos técnicas. La dispersión que tiene cada uno de estos estimadores se la puede observar en los gráficos de las varianzas que se muestran en la figura 4.9.

FIGURA 4.9 Varianza de los Estimadores de la Media



A continuación se presenta un intervalo del 95% de confianza para la media poblacional, para cada una de las técnicas de muestreo.

De lo que obtenemos un límite superior y uno inferior, a continuación se presentan los intervalos obtenidos.

Intervalo para la Media del Muestreo Aleatorio Simple

Límite Superior	4,72520
Límite Inferior	4,543 16
Desviación Estándar	0,45871

Intervalo para la Media del Muestreo Estratificado

Límite Superior	4,79661
Límite Inferior	4,65260
Desviación Estándar	0,36289

Intervalo para la Media del Muestreo Sistemático.

Límite Superior	4,73705
Límite Inferior	4,58431
Desviación Estándar	0,38488

ESPOL 1959 ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

De estos intervalos podemos observar que el que en promedio, el que tiene menor longitud es el obtenido con el muestreo estratificado al igual que el del muestreo sistemático.

De la misma manera que en los casos anteriores, realizamos la prueba K-S para verificar la distribución de la media; para propusimos las siguientes hipótesis

H_0: la media aritmética tiene una distribución $N(4{,}6677\ ,\ 21{,}1182)$
VS.
H_1: $\neg\ H_0$.

Como se puede observar en la TABLA XXI, el valor **p** de cada una de las pruebas nos permiten afirmar que existe evidencia estadística para aceptar la hipótesis nula antes propuesta.

TABLA XXI. Prueba K-S para las medias $\bar{x}$, $\bar{x}_{est}$ y $\bar{x}_{sist}$.

(Mezcla de Poblaciones: N(4,15) N(2,15), N(8,15), n=100)

Muestreo Aleatorio Simple			
Prueba de Kolmogorov – Smimov usando una distribución Normal(4,67, 2 1,118)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,06482	0,79487
Muestreo Estratificado			
Prueba de Kolmogorov -Smimov usando una distribución Normal (4,67, 2 1,118)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,02332	0,89549
Muestreo Sistemático			
Prueba de Kolmogorov – Smimov usando una distribución Normal(4,67, 2 1,118)			
Variable	N de Casos	Máxima Diferencia	Valor p (2-colas)
MEDIA	100	0,08843	0,4 1474

A continuación presentamos los resultados obtenidos cuando estimamos el total poblacional con cada una de las técnicas de muestreo.

ESPOL 1958 ESCUELA SUPERIOR POLITECNICA DEL LITORAL CIB - ESPOL

TABLA XXII. Comparación de los resultados con Muestreo Aleatorio Simple, Muestreo Estratificado y Muestreo Sistemático. (Mezcla de Poblaciones: N(4,15) N(2,15), N(8,15), n=100)

	Muestreo Aleatorio Simple	Muestreo Estratificado	Muestreo Sistemático
TOTAL			
Número de Casos	100	100	100
Mínimo	3700,000	403 1,000	3 800,000
Máximo	6200,000	6025,000	6000,000
Mediana	5000,000	5019,500	5050,000
Media	5042,000	5009,230	5082,000
Desviación Estándar	474,0072	425,023 3	484,378 1
Varianza	2,24683E+05	1,80645E+05	2,34622E+05

Al analizar esta TABLA XXII podemos observar que con el muestreo estratificado subestiman al parámetro, mientras que el muestreo aleatorio simple y el sistemático lo sobrestima; con el muestreo estratificado el estimador toma un valor de 5.009 ± 425,02, en el aleatorio simple es igual a 5.042 ± 474,007, y en el sistemático un valor de 5082 ± 484,37

Además de la estadística descriptiva del total para cada método, calculamos el error de estimación obtenido con cada uno de los tipos de muestreo.

Muestreo Aleatorio Simple

$e = |\hat{\theta} - \theta|$

$e = |5042 - 5019|$

$e = 23$

Muestreo Estratificado

$e = |5009 - 5019|$

$e = 10$

Muestreo Sistemático

$e = |5082 - 5019|$

$e = 63$

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB ESPOL

Al analizar estos resultados podemos observar que el error de estimación en el muestreo estratificado es menor que en los otros dos métodos, éste nos da un error en la estimación de 10, mientras que el muestreo aleatorio simple el error es de 23 y en el sistemático el error de 63. En las figuras 4.10, 4.11 y 4.12 se muestran los histogramas de frecuencia para el total obtenido con muestreo aleatorio simple, del muestreo estratificado y del sistemático respectivamente

FIGURA 4.10 Histograma de Frecuencia para el Total

(Muestreo Aleatorio Simple, n=100)

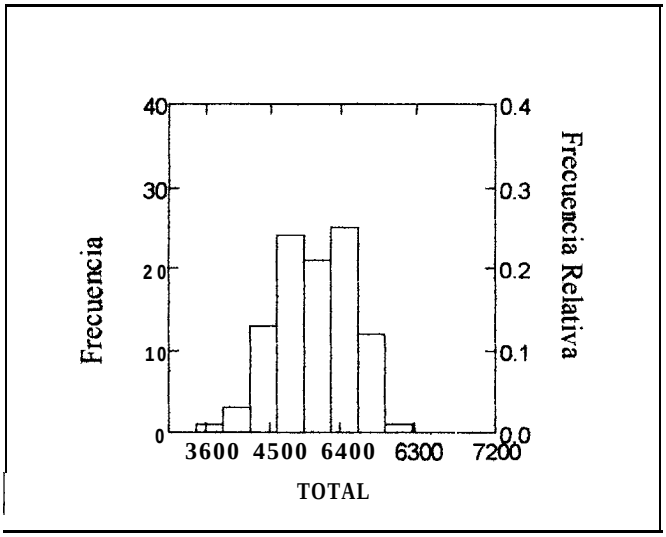


FIGURA 4.11 Histograma de Frecuencia para el Total

(Muestreo Sistemático, n=100)

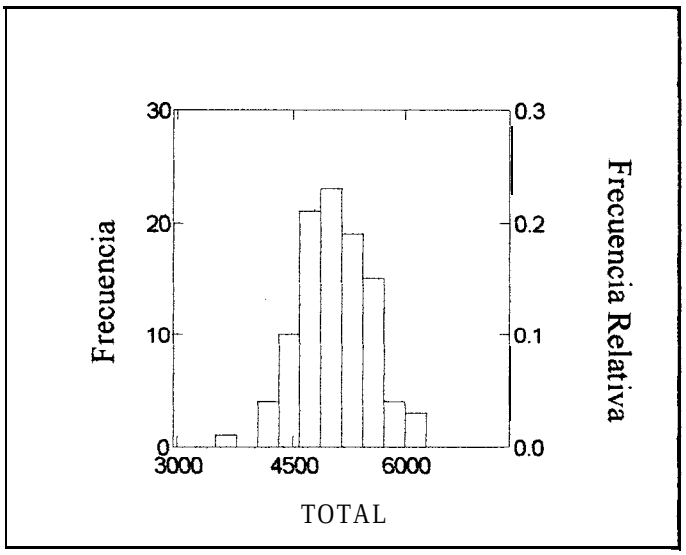
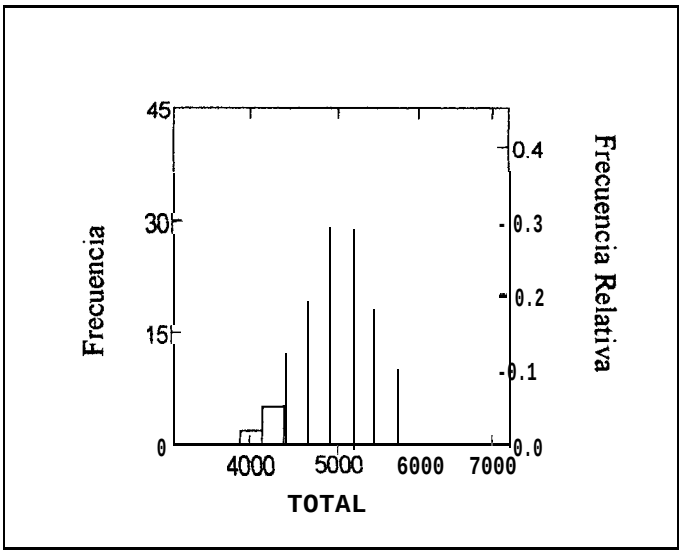


FIGURA 4.12 Histograma de Frecuencia para el Total (Muestreo Estratificado, n=100)



En cada una de las figuras presentadas se puede observar que la dispersión del total es diferente para cada uno de los casos; en el muestreo estratificado el estimador del total tiene una dispersión menor que en los otros dos métodos, podemos observar que en el muestreo aleatorio simple el total se distribuye alrededor de 5042, en el sistemático alrededor de 5082 y en el estratificado esta distribuido alrededor de 5009.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES.

Luego de haber analizado cada uno de los resultados obtenidos en este estudio, podemos concluir lo siguiente:

1. El Muestreo presenta ventajas indiscutibles sobre los censos completos tanto en el aspecto económico como en el técnico, debido a que por medio de esta podemos conocer características de una población utilizando para esto solo una parte de ella, todo esto a bajos costos y con mayor rapidez.

ESPOL
ESCUELA SUPERIOR POLITECNICA DEL LITORAL
CIB - ESPOL

2. El Muestreo Aleatorio Simple, el Muestreo Estratificado y el Muestreo Sistemático, a pesar de ser todas técnicas estadísticas que nos permiten realizar inferencias acerca de una población, podemos observar que el segundo nos proporciona estimadores insesgados de menor varianza; esto nos permite obtener una mayor precisión en la estimación de un parámetro. Además también podemos observar que los estimadores obtenidos a partir del Muestreo

Sistemático, ofrecen una mayor precisión que los que se obtienen de un Muestreo Aleatorio Simple.

3. Los intervalos de confianza construidos para la media de una población mezclada (N(2,10), N(0,30), N(5,8)) a partir de muestras de tamaño n=100 utilizando el muestreo estratificado, tienen una mayor cobertura que los intervalos construidos utilizando el Muestreo Aleatorio Simple o el Muestreo Sistemático; esto nos indica que dichos intervalos de confianza nos proporcionan mayor información a cerca del parámetro estimado. Así tenemos, que en el Muestreo Estratificado la longitud promedio fue de 1,54443 , para el Muestreo Aleatorio Simple fue de 1,7453 y para el Sistemático fue de 1,7400. Es importante también señalar, que al comparar la longitud promedio de los intervalos obtenidos con el muestreo sistemático, con la longitud promedio de los obtenidos a partir de una muestreo aleatorio simple, podemos ver que el primero posee una longitud promedio menor, por lo cual se considera que este mejor que el aleatorio simple.

4. La convergencia que tiene la media aritmética hacia la media poblacional es más lenta cuando la población de la cual se muestrea es discreta (Poisson) y se encuentra mezclada, que si se tratase con una población “pura”, en la cual solo

existiese un tipo de variable aleatoria. Esto se lo puede observar en los valores de p la prueba K-S obtenidos para cada uno de los casos.

5. En general los resultados analíticos no están alejados de los resultados que se obtienen por simulación numérica, sin embargo existen diferencias en cuanto a la eficiencia de los estimadores y al acercamiento hacia el valor esperado.

RECOMENDACIONES.

1. Considerando que el estudio que hemos realizado esta basado en tamaños de muestras menores 0 iguales a 100 , se aconseja realizar simulaciones con muestras de tamaño mayor a este y con mayor cantidad de muestras para verificar si es que los resultados analíticos concuerdan con los resultados numéricos para esos valores.

2. Considerando que en el año 2001 se realizará el censo poblacional, se aconseja repartir un formulario ampliado a una parte de la población, y utilizar una de estas técnicas de muestreo para conocer otras características que pudiera tener la población y que podrían ser de importancia para diseñar proyectos orientados al área social.

Anexos

ANEXO 1

Intervalos del 95% de Confianza

Muestras tomadas de una población Normal (2,10)

Muestreo Aleatorio Simple

6 intervalos no contienen el valor de la media

Muestra	LCI	Media	LCS	S
1	1.35201714	2.32103752	3.29005791	0.49439816
2	0.1844871	1.26656165	2.3486362	0.55207885
3	0.96523692	1.95988073	2.95452454	0.50747133
4	0.59793605	1.82790311	3.05787017	0.62753422
5	1.27220121	2.22276684	3.17333246	0.48498246
6	-0.39671501	1.17308755	2.74289011	0.80091967
7	0.54728087	1.49814629	2.44901171	0.485 13542
8	1.28221366	2.23754691	3.19288016	0.48741492
9	2.21664173	3.13697581	4.05730989	0.46955821
10	2.33746826	3.4057214	4.47397454	0.5450271 1
11	1.0521564	2.53922589	4.02629539	0.75870893
12	0.90210095	2.10501001	3.30791908	0.61372912
13	1.7547038	2.99563064	4.23655748	0.63312594
14	0.98730062	2.1936876	3.40007457	0.61550356
15	1.48898268	2.76395097	4.03891925	0.65049402
16	1.11570982	2.06492273	3.01413563	0.4842923
17	1.31136139	2.53072484	3.75008828	0.62212421
18	1.16457748	2.25924302	3.35390855	0.55850282
19	0.3773936	1.65252111	2.92764862	0.65057526
20	1.12461324	2.3744159	3.62421857	0.63765442
21	0.13812098	1.18792769	2.2377344	0.53561567
22	-0.22403692	0.86208962	1.94821616	0.55414619
23	0.32012246	1.54102305	2.76192364	0.62290846
24	1.35161584	2.45345942	3.555303	0.56216509
25	0.92609829	2.07913825	3.23217822	0.5882857
26	1.557376	2.80687422	4.05637243	0.63749909
27	1.00775684	2.10474162	3.20172641	0.55968612
28	0.93926961	2.1779185	3.41656739	0.63 196372
29	0.6199471	1.9399114	3.25987569	0.67345117
30	1.11833426	1.94908477	2.77983528	0.4238523
31	1.29146564	2.42557528	3.55968492	0.57862737

32	0.92993788	2.10085057	3.27176325	0.59740443
33	0.18428939	1.41329471	2.64230003	0.62704353
34	1.19015374	2.21430475	3.23845576	0.52252602
35	0.47274389	1.48699527	2.50124664	0.51747519
36	0.23649502	1.25152578	2.26655653	0.51787283
37	0'3971872	1.56654049	2.73589377	0.59660882
38	0.37059028	1.52057714	2.670564	0.58672799
39	1.19927185	2.40420581	3.60913977	0.61476222
40	1.64387687	2.85189089	4.05990492	0.61633368
41	0.65213384	1.89307386	3.13401387	0.63313266
42	0.63732231	1.68806066	2.738799	0.53609099
43	1.08918099	2.27878494	3.46838889	0.60694079
44	1.20684278	2.27046066	3.33407854	0.54266218
45	0.65032607	1.57562701	2.50092795	0.47209232
46	1.45462441	2.51285131	3.57107821	0.53991168
47	1.28262758	2.20774308	3.13285858	0.47199771
48	0.54731861	1.556532	2.56574538	0.51490479
49	0.8062229	1.83381434	2.86140578	0.52428135
50	0.97525552	2.28974292	3.60423032	0.67065684
51	0.45674446	1.81677528	3.17680611	0.69389328
52	0.3039181	1.37276063	2.44160317	0.54532782
53	2.56737522	3.64387319	4.72037116	0.54923366
54	1.86841536	2.83820748	3.8079996	0.4947919
55	0.6239217	1.91895518	3.21398866	0.66073 137
56	1.57348713	2.6759645 1	3.7784419	0.56248846
57	0.47354312	1.56939691	2.6652507	0.55910908
58	0.44229824	1.49526494	2.54823 165	0.53722791
59	1.08915827	2.3977313	3.70630434	0.6676393
60	2.15157161	3.03158367	3.91159573	0.44898575
61	1.58962788	2.82018569	4.0507435	0.62783562
62	0.670073 13	1.77034992	2.87062671	0.56136571
63	0.95 185845	2.30963421	3.66740997	0.69274273
64	1.1440828	2.1215321	3.09898141	0.49869863
65	1.88682972	2.74472389	3.60261806	0.437701 ll
66	1.29522437	2.23214865	3.16907293	0.47802259
67	0.65719644	1.61728183	2.57736722	0.48983948
68	0.34595723	1.46736986	2.58878249	0.5721493
69	0.59891072	1.87145422	3.14399771	0.64925688
70	-0.05350717	1.23540877	2.52432471	0.65761017
71	0.5725338	1.98740153	3.40226926	0.72187129
72	1.41631937	2.4683977	3.52047604	0.53677466
73	0.21972654	1.2'7323 166	2.32673678	0.53750261
74	0.58692612	1.70829331	2.8296605 1	0.57212612

75	1.60053088	2.58438911	3.56824735	0.50196849
76	0.93682017	2.00938568	3.08195119	0.5472273
77	0.83281363	1.79651691	2.76022018	0.49168534
78	-0.2432211	0.94382094	2.13086298	0.60563369
79	1.56786743	3.00076007	4.43365272	0.73106768
80	0.9668035 1	1.98759515	3.00838679	0.52081206
81	0.56937382	1.83660499	3.10383617	0.64654652
82	0.32338419	1.65804255	2.99270091	0.68094814
83	1.01081281	1.98906181	2.96731081	0.49910663
84	0.98556283	2.23217124	3.47877964	0.6360247
85	1.18715876	2.34481743	3.50247609	0.59064218
86	0.79892459	1.88899785	2.979071 ll	0.55615983
87	0.63807075	1.72569346	2.81331617	0.55490954
88	-0.07782983	1.06088825	2.19960632	0.58097861
89	0.03754838	1.18200131	2.32645424	0.58390455
90	0.02042062	1.04284548	2.06527033	0.52164533
91	0.95775553	1.97141455	2.98507356	0.51717297
92	1.35302124	2.5158629	3.67870455	0.59328656
93	0.44202808	1.48905276	2.53607744	0.53419627
94	0.16214184	1.2830401	2.40393836	0.57188687
95	-0.04145965	1.01720903	2.07587771	0.54013708
96	0.49440709	1.73944278	2.98447847	0.63522229
97	1.10818136	2.28408166	3.45998197	0.59994913
98	1.17002217	2.23034098	3.2906598	0.54097899
99	1.26433288	2.37230898	3.48028509	0.56529393
100	-0.74039373	0.3049841	1.35036193	0.53335604

Intervalos de Confianza del 95%

Muestras tomadas de una población Normal (2,10)

Muestreo Sistemático

6 intervalos no contienen a la media

Muestra	LCI	Media	LCS	S
1	0.93028159	2.04769177	3.16510195	0.57010724
2	0.49520973	1.4167236	2.33823747	0.47016014
3	-0.01123306	1.26273364	2.53670035	0.64998301
4	2.16660407	3.19130251	4.21600096	0.52280533
5	0.83259015	1.85265296	2.87271578	0.5204402 1
6	1.23965756	2.46158382	3.68351008	0.62343 177
7	0.32758912	1.37791637	2.42824361	0.53588125
8	0.52207499	1.42878615	2.3354973	0.46260773
9	0.89161731	1.89977608	2.90793485	0.51436672
10	0.7527329	1.98048874	3.20824459	0.62640604
11	0.40891649	1.4936608	2.57840512	0.55344098
12	0.88575034	2.20333525	3.52092016	0.6722372
13	0.78647765	1.86839585	2.95031405	0.55 199908
14	0.15591767	1.53391543	2.91191319	0.70306008
15	1.28369301	2.5006038	3.71751459	0.62087285
16	0.93846058	2.38836837	3.83827616	0.73974887
17	1.44825879	2.51624295	3.58422712	0.54488988
18	0.28253412	1.57587083	2.86920754	0.65986567
19	0.85010467	2.06968038	3.2892561	0.6222325 1
20	0.91398928	1.98535835	3.05672743	0.54661687
21	0.15788579	1.27617068	2.39445557	0.57055352
22	0.2752557	1.31875842	2.36226113	0.53239935
23	0.58439662	1.70456637	2.82473612	0.57151518
24	1.03328583	2.27242003	3.51155422	0.63221133
25	1.03520131	2.18659254	3.33798378	0.5874445 1
26	0.20380256	1.29135035	2.37889814	0.55487132
27	1.283773 14	2.41357929	3.54338544	0.57643171
28	0.2655093	1.30180758	2.33810586	0.52872361
29	1.10862439	2.23088128	3.35313817	0.53258005
30	0.13719873	1.44873907	2.7602794	0.66915323
31	1.31525512	2.459533 16	3.60381119	0.58381532
32	1.27263838	2.19819133	3.12374428	0.47222089
33	0.65159838	1.864617	3.07763562	0.61888705
34	1.23156351	2.4433 1106	3.65505861	0.61823855
35	0.70936659	1.75817759	2.80698859	0.53510765
36	0.96738674	2.01806081	3.06873488	0.5360582

37	1.49446637	2.62242768	3.75038899	0.57549047
38	1.30147533	2.37502552	3.44857571	0.54772969
39	0.74850345	1.76175971	2.77501596	0.5 1696748
40	2.01310069	3.14108743	4.26907417	0.57550344
41	2.42703799	3.3122831	4.19752822	0.45165567
42	0.67979875	1.76759374	2.85538873	0.55499745
43	0.23606338	1.34417096	2.45227854	0.56536101
44	0.07821394	1.22666943	2.37512493	0.58594668
45	0.80732935	1.89252544	2.97772152	0.55367147
46	0.17899306	1.36312262	2.54725217	0.60414773
47	-0.12221928	1.03 198339	2.18618606	0.58887891
48	-0.01060006	1.08690764	2.18441534	0.55995291
49	0.49382852	1.62068821	2.7475479	0.57492841
50	1.10312397	2.26227845	3.42143292	0.59140534
51	0.31431679	1.30486717	2.29541755	0.50538285
52	-0.2783577	0.96361654	2.20559078	0.63366033
53	-0.00395737	1.19393305	2.39182346	0.61116858
54	0.87801575	1.93216828	2.9863208 1	0.53783292
55	0.89345765	1.97221534	3.05097302	0.55038658
56	1.03292867	2.08189795	3.13086724	0.53518841
57	1.17250918	2.648614	4.12471881	0.7531147
58	0.7946129	2.07945786	3.36430283	0.65553315
59	0.38735755	1.72301986	3.05868218	0.68 146037
60	0.96204767	1.88369349	2.80533932	0.47022746
61	0.84180937	2.0928946	3.34397983	0.63830879
62	1.14579656	2.21036481	3.27493306	0.54314707
63	1.23152816	2.50171308	3.771898	0.64805353
64	1.61050979	2.76568978	3.92086978	0.58937755
65	-0.1521093	0.78508268	1.72227467	0.47815918
66	0.91337129	2.09815565	3.28294	0.60448181
67	1.01681316	2.25495199	3.49309081	0.63 170348
68	-0.10013035	0.91265926	1.92544887	0.51672939
69	1.33389968	2.3338254	3.33375113	0.51016619
70	1.13523953	2.17618317	3.2171268	0.53 109369
71	0.58127849	1.75352676	2.92577504	0.59808586
72	1.23462003	2.37422371	3.51382739	0.58143045
73	1.04901319	2.24026257	3.43151195	0.60778029
74	1.01887729	2.35510868	3.69134006	0.68175071
75	1.39878428	2.3509364	3.30308853	0.4857919
76	2.23906785	3.37266027	4.50625268	0.57836348
77	0.3577776	1.29031874	2.22285987	0.47578629
78	1.59845549	2.828973 18	4.05949087	0.62781515
79	1.42538983	2.64734731	3.8693048	0.6234477
80	0.25516227	1.33903716	2.52291205	0.5785076

81	1.10793948	1.94377403	2.77960858	0.4264462
82	0.61695376	1.90223508	3.18751641	0.65575578
83	0.83051007	2.06788012	3.30525017	0.63131125
84	0.45662835	1.56238855	2.66814875	0.56416337
85	-0.21710644	1.05638825	2.32988293	0.64974218
86	-0.02215011	1.27110113	2.56435237	0.65982206
87	0.84165966	2.0855 1004	3.32936041	0.63461754
88	1.51851602	2.37166954	3.22482306	0.43528241
89	1.26289367	2.46948187	3.67607007	0.61560622
90	0.41276512	1.47943539	2.54610565	0.54421952
91	-0.0621548	1.02199841	2.10615161	0.55313939
92	0.0785871	1.26747759	2.45636809	0.60657678
93	0.74187662	1.58821634	2.43455606	0.43 180598
94	1.73888383	2.60999094	3.48109806	0.4444424 1
95	0.81451883	1.94456849	3.07461816	0.57655595
96	0.41164153	1.33614977	2.26065801	0.47168788
97	-0.1131642	1.07239406	2.25795232	0.60487666
98	0.93526137	2.07244561	3.20962985	0.58019604
99	1.12598926	2.08024789	3.03450652	0.48686665
100	1.52724093	2.72343962	3.91963831	0.61030545

ANEXO 2

Intervalos del 95% de Confianza

Muestras tomadas de una población mezclada: Normal (2,10),N(0,30) y N(5,8)

Muestreo Aleatorio Simple

13 intervalos no contienen el valor de la media

Muestra	LCI	Media	LCS	S
1	1.85616236	2.63394734	3.41173232	0.39682907
2	1.59249775	2.50038542	3.40827309	0.46320799
3	1.30873652	2.22312906	3.13752161	0.4665268 1
4	0.93025142	1.97596236	3.0216733	0.53352599
5	2.41260049	3.18202832	3.95145615	0.39256522
6	1.32559319	2.21582881	3.10606443	0.45420185
7	1.693317	2.55567578	3.41803456	0.43997897
8	0.6654604	1.56305361	2.46064682	0.45795572
9	1.74392155	2.5765553	3.40918905	0.42481314
10	1.32569412	2.19629435	3.06689458	0.44418379
11	0.55669182	1.49441135	2.43213088	0.47842833
12	1.91339712	2.8169753	3.72055349	0.46100928
13	1.55675644	2.34798202	3.1392076	0.40368652
14	1.86179794	2.75297163	3.64414533	0.45468046
15	1.35798672	2.16617385	2.97436098	0.41234037
16	1.31143246	2.18911845	3.06680444	0.44779897
17	0.72683829	1.63241201	2.53798573	0.46202741
18	1.90801946	2.71958288	3.53114631	0.41406297
19	1.61763324	2.47734239	3.33705155	0.43862712
20	1.63671733	2.58219351	3.52766969	0.48238581
21	2.34805156	3.24662375	4. 14519594	0.4584552
22	0.65874719	1.54040649	2.42206578	0.44982617
23	1.33636689	2.16324511	2.99012334	0.42187664
24	1.30834183	2.15779859	3.00725535	0.4333963
25	1.31858168	2.25434736	3.19011305	0.47743 147
26	1.0232945	1.82483875	2.62638301	0.40895115
27	1.31356761	2.26157784	3.20958806	0.48367868
28	1.40812507	2.24843511	3.08874515	0.42872961
29	1.42152794	2.24245757	3.0633872	0.41884165

30	1.69407454	2.55050756	3.40694057	0.43695562
32	1.92381453	2.79457894	3.66534334	0.44426755
33	2.14932552	3.05352376	3.95772199	0.46132563
34	0.6368502	1.5920011	2.547152	0.48732 189
35	1.79402138	2.59680889	3.3995964	0.40958546
36	2.04187809	2.90085264	3.7598272	$43825232
37	1.42626546	2.28781813	3.14937081	0.43956769
38	1.15941244	2.12833154	3.09725064	0.49434648
39	1.36298782	2.18079916	2.9986105	0.41725068
40	1.09134574	2.00271918	2.91409263	0.46498645
41	0.92408765	1.82068877	2.7172899	0.45744955
42	1.82346209	2.57303964	3.3226172	0.38243753
43	2.01807636	2.89115946	3.76424256	0.44545056
44	1.3782233	2.25789118	3.13755907	0.44881014
45	1.08062771	1.9670678	2.85350789	0.45226535
46	1.23744551	2.1306698	3.02389409	0.45572668
47	1.00326493	1.88281955	2.76237418	0.44875236
48	0.73244863	1.6747235	2.61699836	0.48075248
49	2.51349647	3.48662249	4.45974852	0.49649287
50	2.12460854	2.94991944	3.77523034	0.42107699
51	1.52677861	2.47813719	3.42949576	0.48538703
52	2.49671619	3.35664245	4.21656871	0.43873789
53	1.63613889	2.48896211	3.34178533	0.43511389
54	1.87844138	2.8039018	3.72936223	0.472 17369
55	1.74929709	2.59956068	3.44982426	0.43380795
56	1.52632013	2.31709617	3.10787221	0.40345716
57	1.209455 15	2.14362534	3.07779553	0.47661744
58	1.17390825	2.03668616	2.89946408	0.44019281
59	2.04415495	2.8992456	3.75433625	0.43627074
60	1.73287432	2.5181567	3.30343908	0.40065428
61	1.52551682	2.36823247	3.21094813	0.42995697
62	1.02426546	1.87179258	2.7193197	0.4324118
63	1.41239078	2.36234843	3.31230609	0.48467227
64	1.2394079	2.11231016	2.98521242	0.4453583
65	0.50066006	1.40818522	2.31571038	0.46302304
66	1.40976725	2.2047006	2.99963395	0.40557824
67	1.41711256	2.24512227	3.07313198	0.42245393
68	1.60755085	2.45910078	3.31065071	0.43446425
69	1.11064443	1.97522917	2.83981392	0.441 11467
70	1.36628717	2.22750892	3.08873067	0.43939885
71	1.36699676	2.11929208	2.8715874	0.38382414
72	1.25661716	2.09431447	2.93201178	0.42739659
73	1.62497052	2.38799336	3.15101621	0.38929737

74	1.75759303	2.53357055	3.30954808	0.3959069
75	1.16630174	2.09770487	3.029108	0.47520568
76	1.84726966	2.70671353	3.56615739	0.43849177
77	1.54258132	2.440351	3.33812067	0.45804575
78	1.76781007	2.63339567	3.49898126	0.44 16253
79	2.21800897	3.04940807	3.88080717	0.42418321
80	1.29161821	2.16504483	3.03847145	0.44562583
81	0.84350739	1.80878322	2.77405904	0.49248766
82	1.73764992	2.66438026	3.5911106	0.47282 16
83	1.7283301	2.60763573	3.48694136	0.44862532
84	1.92735169	2.78401955	3.64068742	0.43707544
85	1.70381233	2.52263912	3.34146591	0.41776877
86	1.50057797	2.29895532	3.09733267	0.40733538
87	1.49327896	2.44921604	3.40515312	0.487723
88	1.14785899	2.08791808	3.02797717	0.47962198
89	0.94668008	1.96390432	2.98 112856	0.51899196
90	2.23005427	3.02707092	3.82408756	0.40664115
91	1.58166265	2.43560404	3.28954543	0.43568438
92	1.35850922	2.29810959	3.23770996	0.47938794
93	1.61201911	2.5370961	3.46217309	0.47197806
94	1.3864734 1	2.32039943	3.25432545	0.47649287
95	1.09953777	1.94794736	2.79635696	0.43286204
96	2.50384635	3.28424716	4.06464798	0.39816368
98	1.3 1370002	2.28329161	3.25288321	0.49468959
99	1.4786285	2.289279	3.0999295	0.41359719
	2.07623921	2.85375668	3.63127416	0.39669259
100	1.36563063	2.35630785	3.34698506	0.50544756

Intervalos del 95% de Confianza

Muestras tomadas de una población mezclada: Normal (2,10),N(0,30) y N(5,8)

Muestreo Estratificado

10 intervalos no contienen el valor de la media

Muestra	LCI	Media	LCS	S
1	1.34569199	2.09003174	2.83437149	0.379765 18
2	1.62126574	2.31718024	3.01309473	0.35505842
3	1.84901823	2.69586449	3.54271075	0.43206442
4	1.78404447	2.54141256	3.29878065	0.38641229
5	2.24226424	2.98349754	3.72473084	0.378 X8025
6	1.31377216	1.957761	2.60174984	0.32856573
7	1.54608983	2.30881281	3.07153579	0.38914438
8	1.1394171	1.94033658	2.74125605	0.40863239
9	0.98084016	1.78849083	2.5961415	0.41206667
10	0.47732374	1.33762237	2.197921	0.43892787
11	1.2766783 1	2.01967954	2.76268078	0.37908226
12	1.27129061	2.02978153	2.78827246	0.38698517
13	1.98189654	2.78623237	3.5905682	0.41037542
14	0.67417534	1.45210897	2.23004259	0.39690491
15	1.32405825	2.1422706	2.96048295	0.41745528
16	1.71117647	2.40197858	3.09278068	0.35245005
17	1.7350352	2.58922107	3.44340694	0.43580912
18	1.81240803	2.56226562	3.3121232	0.3825804
19	1.94589948	2.64648645	3.34707341	0.35744233
20	1.28041885	2.12776578	2.97511272	0.43231986
21	0.90288339	1.63999526	2.37710713	0.37607749
22	1.4910533	2.24872413	3.00639496	0.38656675
23	1.16341942	1.84682124	2.53022306	0.3486744
24	1.80419407	2.47677344	3.14935281	0.343 15274
25	1.19827225	1.94899918	2.6997261	0.38302394
26	1.03147992	1.7358051	2.44013028	0.35934958
27	1.10670138	1.87908643	2.65 147148	0.39407401
28	1.37976244	2.086974	2.79418555	0.36082222
29	1.52120032	2.34087778	3.16055523	0.41820278
30	1.69235788	2.50180112	3.31124435	0.41298124
31	1.61647746	2.42439189	3.23230632	0.41220124
32	0.66179499	1.42963367	2.19747234	0.39175443
33	1.65410061	2.38848787	3.12287514	0.37468738
34	1.31866261	2.18293604	3.04720946	0.44095583

35	1.69583327	2.55568466	3.41553605	0.43869969
36	1.97701158	2.83619967	3.69538776	0.43836127
37	1.60902699	2.50403988	3.39905277	0.45663923
38	1.67572194	2.36318063	3.05063932	0.35074423
39	1.91541714	2.78587296	3.65632878	0.44411011
40	1.84088126	2.63364262	3.42640398	0.40447008
41	1.72372671	2.4907802	3.25783369	0.39135382
42	1.60141287	2.32209351	3.04277415	0.3676942
43	1.51678608	2.34675603	3.17672598	0.42345406
44	2.24518253	3.13478426	4.02438599	0.45387843
45	0.76890879	1.51923087	2.26955295	0.38281739
46	2.0614551	2.7984311	3.53540711	0.37600816
47	1.31552466	2.048894	2.78226333	0.37416803
48	1.32984204	1.97826438	2.62668671	0.33082772
49	2.05114503	2.87655026	3.70195549	0.42112512
50	1.62816208	2.40021088	3.17225968	0.39390245
51	0.96891329	1.84879855	2.72868381	0.44892105
52	1.74874354	2.46141031	3.17407708	0.3636055
53	1.59782925	2.46232677	3.32682428	0.44107016
54	1.43501989	2.10373157	2.77244325	0.34117943
55	1.15120234	1.8720861	2.59296986	0.36779784
56	1.55292625	2.32814336	3.10336046	0.3955 1893
57	1.54201	2.4229238	3.3038376	0.44944582
58	2.09665954	2.7889794	3.48129926	0.35322442
59	2.17854781	2.93555981	3.69257181	0.38623061
60	1.6390387	2.38955527	3.14007184	0.38291662
61	1.03380571	1.79841394	2.56302218	0.39010624
62	2.24792708	2.9006415	3.55335593	0.33301756
63	1.83652428	2.52754728	3.21857029	0.35256276
64	1.67585759	2.50600897	3.33616034	0.42354662
65	1.344991	2.02519209	2.705393 19	0.34704138
66	1.20470528	1.98474814	2.764791	0.39798105
67	1.77826996	2.4521957	3.12612145	0.34383966
68	1.74293298	2.45256587	3.16219876	0.3620576
69	1.82872114	2.60626884	3.38381655	0.39670801
70	1.40262594	2.16801394	2.93340194	0.39050408
71	1.74988066	2.515'15539	3.28043012	0.39044629
72	1.31019107	2.06044661	2.81070215	0.38278344
73	1.28150415	2.05022328	2.81894241	0.39220364
74	1.38274467	2.27730366	3.17186265	0.45640765
75	1.40028634	2.18234776	2.96440918	0.39901093
76	1.67868772	2.48485089	3.29101405	0.41130774
77	0.73349493	1.4083363 1	2.0831777	0.34430683

78	1.34134966	2.24849661	3.15564355	0.46283007
79	2.52250313	3.1993819	3.87626068	13.34534631
80	1.21633203	2.00833478	2.80033753	0.40408304
81	2.09298399	2.95605799	3.81913198	0.44034387
82	1.68259047	2.51417665	3.34576283	0.42427866
83	1.6658605 1	2.41571001	3.16555952	0.38257628
84	1.53321613	2.31925814	3.10530015	0.40104184
85	2.01925914	2.84778775	3.6763 1635	0.42271868
86	1.29406997	2.11089272	2.92771547	0.4167463
87	0.87488315	1.63993332	2.40498348	0.39033 172
88	1.79096266	2.62355504	3.45614743	0.42479203
89	1.96236245	2.72219819	3.48203393	0.3876713
90	1.51487826	2.31151207	3.10814589	0.40644582
91	1.72941206	2.60540097	3.48138987	0.446933 ll
92	0.91599668	1.68474188	2.45348708	0.39221694
93	1.71212679	2.47165197	3.23117716	0.3875 1285
94	1.73746279	2.38994375	3.04242472	0.33289845
95	1.64869998	2.47444148	3.30018298	0.42129668
96	1.16642431	1.9122 1695	2.65780959	0.38045543
97	1.78319488	2.505 18865	3.22718243	0.36836417
98	1.3048826	2.10005965	2.8952367	0.40570258
99	1.83323926	2.62856779	3.42389633	0.40577986
100	0.94918127	1.71473356	2.48028584	0.3905879

Intervalos del 95% de Confianza

Muestras tomadas de una población mezclada: Normal (2,10),N(0,30) y N(5,8)

Muestreo Sistemático

ll intervalos no contienen el valor de la media

Muestra	LCI	Media	LCS	
1	1.34646249	2.41015769	3.47385289	0.54270163
2	2.02205451	2.9114159	3.80077729	0.45375581
3	2.39898071	3.13538557	3.87179043	0.37571677
4	1.69040584	2.51698906	3.34357228	0.42172613
5	1.20826182	2.1094909	3.01071998	0.45981075
6	1.24190038	2.11953725	2.99717412	0.44777391
7	0.70497018	1.67423561	2.64350104	0.494523 18
8	1.51068283	2.39543045	3.28017806	0.45140185
9	0.93924291	1.78882617	2.63840944	0.43346085
10	2.15469332	3.00632314	3.85795296	0.43450501
11	0.90529173	1.77584994	2.64640815	0.44416235
12	1.76329542	2.69013532	3.61697522	0.4728775
13	2.15469332	3.00632314	3.85795296	0.43450501
14	1.21632201	2.10554022	2.99475843	0.45368276
15	1.92964532	2.7134973	3.49734928	0.39992448
16	1.23235203	2.1943001	3.15624816	0.49078983
17	1.01341848	1.92664635	2.83987422	0.46593259
18	0.90529173	1.77584994	2.64640815	0.44416235
19	1.8448127	2.69393395	3.54305519	0.433225 13
20	1.32571899	2.15298717	2.98025535	0.4220756
21	1.12364459	1.97313503	2.82262548	0.43341349
22	1.5410051	2.3795788	3.2181525	0.42784372
23	1.7933304	2.69643689	3.59954338	0.46076862
24	1.19065911	2.04396508	2.89727105	0.43536019
25	1.21632201	2.10554022	2.99475843	0.45368276
26	1.38447249	2.18805531	2.99163813	0.40999123
27	1.22217449	2.15205894	3.08194338	0.47443084
28	1.92964532	2.7134973	3.49734928	0.39992448
29	1.73918981	2.68070881	3.62222781	0.48036684
30	1.08591468	2.0908842	3.09585372	0.51273955
31	1.53902884	2.39345208	3.24787532	0.43593022
32	2.18936431	3.03258134	3.87579837	0.43021277
33	0.90529173	1.77584994	2.64640815	0.44416235
34	1.42013562	2.21862991	3.0171242	0.407395051

35	1.36966357	2.24594097	3.12221837	0.4470803 1
36	1.93234685	2.85807086	3.78379488	0.47230817
37	1.42013562	2.21862991	3.0171242	0.40739505
38	0.90529173	1.77584994	2.64640815	0.44416235
39	1.38447249	2.18805531	2.99163813	0.40999123
40	1.24919318	2.14009552	3.03099786	0.45454201
41	0.91084723	1.82511832	2.7393894	0.46646484
42	1.690405 84	2.51698906	3.34357228	0.42172613
43	1.60048338	2.42027491	3.24006643	0.41826098
44	2.15469332	3.00632314	3.85795296	0.43450501
45	2.28324435	3.09224217	3.90123999	0.41275399
46	1.410222	2.2493234	3.08842479	0.42811296
47	1.73918981	2.68070881	3.62222781	0.48036684
48	1.53902884	2.39345208	3.24787532	0.43593022
49	1.5410051	2.3795788	3.2181525	0.42784372
50	1.93234685	2.85807086	3.78379488	0.472308 17
51	1.23235203	2.1943001	3.15624816	0.49078983
52	1.53902884	2.39345208	3.24787532	0.43593022
53	1.93234685	2.85807086	3.78379488	0.472308 17
54	1.38447249	2.18805531	2.99163813	0.40999123
55	1.87283833	2.75324354	3.63364876	0.44918633
56	2.39898071	3.13538557	3.87179043	0.37571677
57	0.70497018	1.67423561	2.64350104	0.49452318
58	1.83538785	2.70633128	3.57727472	0.4443589
59	2.18936431	‘3.03258234	3.87579837	0.43021277
60	1.95585148	2.89351499	3.8311785	0.47839975
61	1.1076532	1.96549969	2.82334618	0.43767678
62	1.80554683	2.76787044	3.73019404	0.49098143
63	1.63645354	2.48559269	3.33473 183	0.43323426
64	1.21632201	2.10554022	2.99475843	0.45368276
65	1.276602 18	2.17891816	3.08123414	0.4603653
66	1.22514363	2.09229651	2.95944938	0.44242494
67	1.09785706	1.99112556	2.88439405	0.45574923
68	1.20191792	1.96784371	2.7337695	0.39077846
69	1.2280953	2.00718549	2.78627569	0.397495
70	1.42728304	2.36684125	3.30639946	0.47936643
71	1.59217198	2.46468948	3.33720699	0.44516199
72	1.83813421	2.68032895	3.52252368	0.42969119
73	1.80554683	2.76787044	3.73019404	0.49098 143
74	1.12021126	1.94469125	2.76917125	0.42065306
75	1.69040584	2.51698906	3.34357228	0.42172613
76	1.93234685	2.85807086	3.78379488	0.47230817
77	1.79972126	2.65042656	3.50113185	0.4340333 1
78	1.50609541	2.35971596	3.21333651	0.43552069

79	1.23235203	2.1943001	3.15624816	0.49078983
80	1.50609541	2.35971596	3.21333651	0.43552069
81	2.30908071	3.13528557	3.87170043	0.37571677
82	1.38447249	2.18805531	2.99163813	0.40999123
83	1.59217198	2.46468948	3.33720699	0.44516199
84	1.2280953	2.00718549	2.78627569	0.397495
85	1.93234685	2.85807086	3.78379488	0.47230817
86	1.55533595	2.45229011	3.34924426	0.45762967
87	0.91084723	1.82511832	2.7393894	0.46646484
88	0.70102114	1.57062231	2.44022348	0.44367407
89	0.93924291	1.78882617	2.63840944	0.43346085
90	1.36966357	2.24594097	3.12221837	0.4470803 1
91	1.20191792	1.9678437 1	2.7337695	0.39077846
92	0.90089319	1.81788354	2.73487389	0.46785222
93	2.28324435	3.09224217	3.90123999	0.41275399
94	0.95110444	1.82066866	2.69023287	0.44365521
95	1.23235203	2.1943001	3.15624816	0.49078983
96	1.97940123	2.83326696	3.6871327	0.43564578
97	1.69040584	2.51698906	3.34357228	0.42172613
98	1.22974827	2.07567373	2.92159919	0.43159462
99	1.08591468	2.0908842	3.09585372	0.51273955
100	1.31898778	2.13660999	2.9542322	0.41715419

BIBLIOGRAFIA

1. John E. Freund / Ronald E. Walpole, Estadística Matemática con Aplicaciones, Prentice- Hall Hispanoamericana, S.A Cuarta Edición , 1990.

2. Willian Cochran, Técnicas de Muestreo, Editorial Continental, S.A, 1990

3. Sheaffer McClave, Probabilidad y Estadística para Ingeniería, Editorial Iberoamérica, 1993.

4. Francisco Azorín / José Luis Sánchez Crespo, Métodos y Aplicaciones del Muestreo, Alianza Editorial S.A, 1986.

5. Internet: http:// irc.unb.ca/web/transpo/mytnet/....

6. Internet: http://www.worksafe/fulltext/docs/h4/501 .htm

7. Internet: http:// www.stat.ucla.edu

8. Internet: http: drott.cis.drexel.edu/sample/. . .