Universidad Autónoma del Estado de México
Facultad de Economía

Lic. En Actuaria

Unidad de Aprendizaje: Muestreo

Fundamentos de Muestreo (2)

Muestreo Aleatorio Estratificado (MAE)
Muestreo Aleatorio por Conglomerados (MAC)

M. En C. Rafael Morales Ibarra

Febrero 2019

Contenido

Guión explicativo de uso del material

Este conjunto de plantillas es la continuidad del material en diapositivas que conforma el archivo de elementos de muestreo Primera parte, en el cual se desarrollo la parte conceptual del muestreo, así como del diseño de las técnicas de muestreo aleatorio simple y el muestreo sistemático.

En esta segunda parte, se analiza al **muestreo aleatorio estratificado** (MAE) la cual es una herramienta propia del muestreo tipo probabilístico en la cual el investigador separa a toda la población en diversas subcategorías conocidas como estratos, para posteriormente, escoge al azar a los individuos finales de los variados estratos de manera equilibrada.

Posteriormente se aborda al **muestreo por conglomerados** (MAC), una técnica aleatoria, con la cual en lugar de considerar cada elemento de la población, lo que consideramos son "conglomerados de elementos". Se analiza también el proceso para elegir aleatoriamente uno o varios conglomerados así la muestra estará formada por todos los elementos de dichos conglomerados.

La lógica de este material es que estas plantillas buscan ser el complemento a las "*Notas para acompañar el curso de muestreo"* pero de una forma más puntual y buscando suplir los formularios que son necesarios en los cursos de muestreo, esto sin dejar el rigor teórico con el que se trata el curso.

Se incorporan una serie de ejemplo para ilustrar la aplicación, y se integra la bibliografía en la que alumno podría abundar tanto en la teoría como en la aplicación a casos reales.

Programa por competencias de la UA: Muestreo

Universidad Autónoma del Estado de México
Secretaría de Docencia
Coordinación General de Estudios Superiores
Programa Institucional de Innovación Curricular

Programa de Estudios por Competencias

I. IDENTIFICACIÓN DE LA UNIDAD DE APRENDIZAJE.

ORGANISMO ACADÉMICO

FACULTAD DE ECONOMÍA

Programa Educativo: Licenciatura en Actuaría | Área de docencia: Métodos Matemáticos y Estadísticos

Aprobación por los H.H. Consejos Académico y de Gobierno | Fecha: 1 de febrero de 2007 | Programa elaborado por: Juan José Lechuga Arizmendi, Ricardo Rodríguez Marcial, Jesús Salgado Vega, Daniel Hadad Cartas | Fech 10 d

Nombre de la Unidad de aprendizaje: MUESTREO

Clave	Horas de teoría	Horas de práctica	Total de horas	Créditos	Tipo de Unidad de Aprendizaje	Carácter de la Unidad de Aprendizaje
L43225	4	2	6	10	Curso	Obligatorio

Prerrequisitos (Conocimientos Previos): Inferencia Estadística

Unidad de Aprendizaje Antecedente

Unidad de A... Cálculo Actu...

Programas educativos en los que se imparte: Licenciatura en Actuaría

Universidad Autónoma del Estado de México
Secretaría de Docencia
Coordinación General de Estudios Superiores
Programa Institucional de Innovación Curricular

II. PRESENTACIÓN

La unidad de aprendizaje busca capacitar al alumno en el estudio de muestreo

III. LINEAMIENTOS DE LA UNIDAD DE APRENDIZAJE

DOCENTE	DISCENTE
El profesor deberá cubrir la totalidad de los temas del curso y además debe tener conocimiento de todos los temas, deberá asistir a la totalidad de las clases	El alumno resolverá problemas fundamentales

IV. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE

Tanto el profesor como el alumno están comprometidos a llevar un curso según los lineamientos que determine la institución.

El alumno deberá comprender los conceptos y utilizar la investigación de operaciones determinística para su formación como

V. COMPETENCIAS GENÉRICAS

Aplicar el muestreo a problemas en el área de la actuaría

VI. ÁMBITOS DE DESEMPEÑO PROFESIONAL

Salón de clase

Sala de computo

Universidad Autónoma del Estado de México
Secretaría de Docencia
Coordinación General de Estudios Superiores
Programa Institucional de Innovación Curricular

VII. ESCENARIOS DE APRENDIZAJE

Sustantivo

VIII. NATURALEZA DE LA COMPETENCIA
(Inicial, entrenamiento, complejidad creciente, ámbito diferenciado)

Competencia inicial

IX. ESTRUCTURA DE LA UNIDAD DE APRENDIZAJE

Unidad 1. Planeación de la encuesta por muestreo
Unidad 2. Diseño del cuestionario
Unidad 3. Diseño de la muestra
Unidad 4. Tamaño de la muestra
Unidad 5. Muestreo aleatorio simple
Unidad 6. Muestreo estratificado
Unidad 7. Muestreo sistemático
Unidad 8. Muestreo por conglomerados

Universidad Autónoma del Estado de México
Secretaría de Docencia
Coordinación General de Estudios Superiores
Programa Institucional de Innovación Curricular

X.- SECUENCIA DIDÁCTICA

UNIDAD 1. PLANEACIÓN DE LA ENCUESTA POR MUESTREO

1.1 Puntos básicos de la planeación de la encuesta
1.2 Objetivos de la encuesta
1.3 Tamaño de la población
1.4 Sesgo en el diseño de la encuesta

UNIDAD 2. DISEÑO DEL CUESTIONARIO

2.1 Objetivos y limitaciones
2.2 Tipo de cuestionario a emplear
2.3 Formulación de la encuesta
2.4 Zona de estudio
2.5 Recolección de información
2.6 Prueba piloto
2.7 Errores en la aplicación
2.8 Corrección del cuestionario
2. 9 Aplicación
2.10 Codificación

UNIDAD 3. DISEÑO Y TAMAÑO DE LA MUESTRA

3.1 Tamaño de la muestra

3.2 Error en la muestra

3.3 Diseño y tamaño de la muestra por proporciones

3.4 Diseño y tamaño de la muestra por medias

UNIDAD 4. MUESTREO ALEATORIO SIMPLE

4.1 Principios básicos

4.2 Definición de muestreo aleatorio simple

4.3 Diseño de muestra por muestreo aleatorio simple

4.4 Problemas

UNIDAD 5. MUESTREO ESTRATIFICADO

5.1 Principios básicos

5.2 Definición de muestreo estratificado

5.3 Diseño de muestra por muestreo estratificado

5.4 Problemas

UNIDAD 6. MUESTREO SISTEMÁTICO

6.1 Principios básicos

6.2 Definición de muestreo sistemático

6.3 Diseño de muestra por muestreo sistemático

6.4 Problemas

UNIDAD 7. MUESTREO POR CONGLOMERADOS

7.1 Principios básicos

7.2 Definición de muestra por muestreo por conglomerados

7.3 Diseño de muestra por muestreo por conglomerados

7.4 Problemas

XI. DESARROLLO DE LA UNIDAD DE APRENDIZAJE

UNIDAD DE COMPETENCIA I	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valor
Unidad I. Planeación de la encuesta por muestreo	Conocer conceptos y antecedentes de planeación de la encuesta por muestreo	Conocer definición de investigación de operaciones	Trabajo en equipo -razonamiento matemático -responsabilidad en el trabajo

ESTRATEGIAS DIDÁCTICAS:	RECURSOS REQUERIDOS	TIEMPO DESTINADO
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	-pizarrón -proyector de acetatos Computadora	4 horas teóricas y 2 horas practicas

CRITERIOS DE DESEMPEÑO I	EVIDENCIAS	
	DESEMPEÑO	PRODUCTOS
1.1. Puntos básicos de la planeación de la encuesta	Se refiere a que cada alumno conozca y comprenda cada antecedente	Conocimiento y manejo correcto de los conc

1.2. Objetivos de la encuesta	Se refiere a que cada alumno conozca y comprenda cada definición del tema	Conocimiento y manejo correcto de los
1.3. Tamaño de la población	Se refiere a que cada alumno conozca y comprenda cada definición del tema	Conocimiento y manejo correcto de los
1.4. Sesgo en el diseño de la encuesta	Se refiere a que cada alumno conozca y comprenda cada definición del tema	Conocimiento y manejo correcto de los

UNIDAD DE COMPETENCIA II	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Act
UNIDAD II Diseño del cuestionario	Conocer conceptos básicos de diseño de cuestionario, recolección de datos, codificación, formulación, prueba piloto, codificación	Conocer definición de	Trabajo en equipo -razonamiento ma -responsabilidad e

ESTRATEGIAS DIDÁCTICAS:	RECURSOS REQUERIDOS	TIEMPO DESTINADO
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	-pizarrón -proyector de acetatos Computadora	4 horas teóricas y 2 horas practicas

CRITERIOS DE DESEMPEÑO II	EVIDENCIAS	
	DESEMPEÑO	PR
2.1 Objetivos Y Limitaciones	El alumno resolverá problemas del tema	Que el alumno apren tema
2.2 Tipo De Cuestionario A Emplear	El alumno resolverá problemas del tema	Que el alumno apren tema
2.3 Formulación De La Encuesta	El alumno resolverá problemas del tema	Que el alumno apren tema
2.4 Zona De Estudio	El alumno resolverá problemas del tema	Que el alumno apren tema

2.5 Recolección De Información	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.6 Prueba Piloto	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.7 Errores En La Aplicación	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.8 Corrección Del Cuestionario	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2. 9 Aplicación	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.10 Codificación	El alumno resolverá problemas del tema	Que el alumno aprenda tema
2.11 Interpretación Y Resultados Del Cuestionario	El alumno resolverá problemas del tema	Que el alumno aprenda tema

UNIDAD DE COMPETENCIA III	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitu
UNIDAD III Diseño y tamaño de la muestra	Conceptos básicos de diseño y tamaño de la muestra	Manejar, resolver problemas de diseño y tamaño de la muestra	Trabajo en equipo -razonamiento matem -responsabilidad en e

ESTRATEGIAS DIDÁCTICAS:	RECURSOS REQUERIDOS	TIEMPO DESTINADO
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	-pizarrón -proyector de acetatos Computadora	4 horas teóricas y 2 horas practicas

CRITERIOS DE DESEMPEÑO III	EVIDENCIAS	
	DESEMPEÑO	PROD
3.1 Tamaño De La Muestra	El alumno resolverá ejemplos del tema	Conocimiento y manejo estudiados
3.2 Error En La Muestra	El alumno resolverá ejemplos del tema	Conocimiento y manejo estudiados

3.3 Diseño Y Tamaño De La Muestra Por Proporciones	El alumno resolverá ejemplos del tema	Conocimiento y estudiados
3.4 Diseño Y Tamaño De La Muestra Por Medias	El alumno resolverá ejemplos del tema	Conocimiento y estudiados

UNIDAD DE COMPETENCIA IV	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	
UNIDAD IV Muestreo aleatorio simple	Conceptos básicos de muestreo aleatorio simple	Manejar y resolver problemas de muestreo aleatorio simple	Trabajo en e -razonamien -responsabil

ESTRATEGIAS DIDÁCTICAS:	RECURSOS REQUERIDOS	TIEMPO DESTI
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	-pizarrón -proyector de acetatos Computadora	4 horas teóricas 2 horas practicas

CRITERIOS DE DESEMPEÑO IV	EVIDENCIAS	
	DESEMPEÑO	
4.1 Principios Básicos	El alumno resolverá ejercicios del tema	Conocimiento y estudiados
4.2 Definición De Muestreo Aleatorio Simple	El alumno resolverá ejercicios del tema	Conocimiento y estudiados
4.3 Diseño De Muestra Por Muestreo Aleatorio Simple	El alumno resolverá ejercicios del tema	Conocimiento y estudiados
4.4 Problemas	El alumno resolverá ejercicios del tema	Conocimiento y estudiados

UNIDAD DE COMPETENCIA V	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	
UNIDAD V. Muestreo estratificado	Conceptos básicos de muestreo estratificado	Manejar y resolver problemas de muestreo estratificado	Trabajo en eq -razonamient -responsabilid

ESTRATEGIAS DIDÁCTICAS:	RECURSOS REQUERIDOS	TIEMPO DESTIN
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	-pizarrón -proyector de acetatos Computadora	4 horas teóricas y 2 horas practicas

CRITERIOS DE DESEMPEÑO V	EVIDENCIAS	
	DESEMPEÑO	
5.1 Principios Básicos	El alumno resolverá ejercicios del tema	El alumno deberá
5.2 Definición De Muestreo Estratificado	El alumno resolverá ejercicios del tema	El alumno deberá
5.3 Diseño De Muestra Por Muestreo Estratificado	El alumno resolverá ejercicios del tema	El alumno deberá
5.4 Problemas	El alumno resolverá ejercicios del tema	El alumno deberá

UNIDAD DE COMPETENCIA VI	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	
UNIDAD VI. Muestreo sistemático	Conceptos básicos de muestreo sistemático	Manejar y resolver problemas de muestreo sistemático	Trabajo en eq -razonamient -responsabilid

ESTRATEGIAS DIDÁCTICAS:	RECURSOS REQUERIDOS	TIEMPO DESTIN
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	-pizarrón -proyector de acetatos Computadora	4 horas teóricas y 2 horas practicas

CRITERIOS DE DESEMPEÑO VI	EVIDENCIAS	
	DESEMPEÑO	
6.1 Principios Básicos	El alumno resolverá ejercicios del tema	El alumn
6.2 Definición De Muestreo Sistemático	El alumno resolverá ejercicios del tema	El alumn
6.3 Diseño De Muestra Por Muestreo Sistemático	El alumno resolverá ejercicios del tema	El alumn
6.4 Problemas	El alumno resolverá ejercicios del tema	El alumn

UNIDAD DE COMPETENCIA VII	ELEMENTOS DE COMPETEN		
	Conocimientos	Habilidades	
UNIDAD VII. Muestreo por conglomerados	Conceptos básicos de	Manejar y resolver problemas de	Trab reso resp

ESTRATEGIAS DIDÁCTICAS	RECURSOS REQUERIDOS	TIEMPO
Demostración con práctica, elaboración y solución de problemas en clase y extra clase.	pizarrón proyector de acetatos Computadora	4 horas t 2 horas p

CRITERIOS DE DESEMPEÑO VII	EVIDENCIAS	
	DESEMPEÑO	
7.1 Principios Básicos	El alumno resolverá ejercicios del tema	El alumn
7.2 Definición De Muestreo Por Conglomerados	El alumno resolverá ejercicios del tema	El alumn
7.3 Diseño De Muestra Por Muestreo Por Conglomerados	El alumno resolverá ejercicios del tema	El alumn
7.4 Problemas	El alumno resolverá ejercicios del tema	El alumn

XII. EVALUACIÓN Y ACREDITACIÓN

La evaluación para esta unidad de aprendizaje es cumplir con el 80% de asistencias para poder presentar los exámen de suficiencia, aprobándolos con calificación mayor a seis puntos.

XIII. REFERENCIAS

1. Cochran W. "Técnicas De Muestreo". Editorial Cecsa, 14 Edición, 1998, México, D. F.
2. Liningere, Ch. "La Encuesta Por Muestreo: Teoría Y Práctica". Editorial Cecsa, 1ra Edición, 1984, México, D. F.
3. Mendenhall, W. "Elementos De Muestreo". Grupo Editorial Iberoamérica, 1987, México, D. F.
4. Des, Raj. "Teoría Del Muestreo". Fondo De Cultura Económica, 1995, México, D. F.
5. Lohr, S. "Muestreo: Diseño Y Análisis". International Thomson Editores, 2000, México, D. F.
6. Holguín, F, Hayashi L. "Elementos De Muestreo Y Correlación". Textos Universitarios, Segunda Edición, 1977, México,
7. Krueckeberg, D. "Análisis D E Planificación Urbana Métodos Y Modelos". Editorial Limusa, Primera Reimpresión, 1994
8. Walpole, R. "Probabilidad Y Estadística Para Ingenieros". Editorial Prentice Hall, Sexta Edición, 1998, México, D. F.

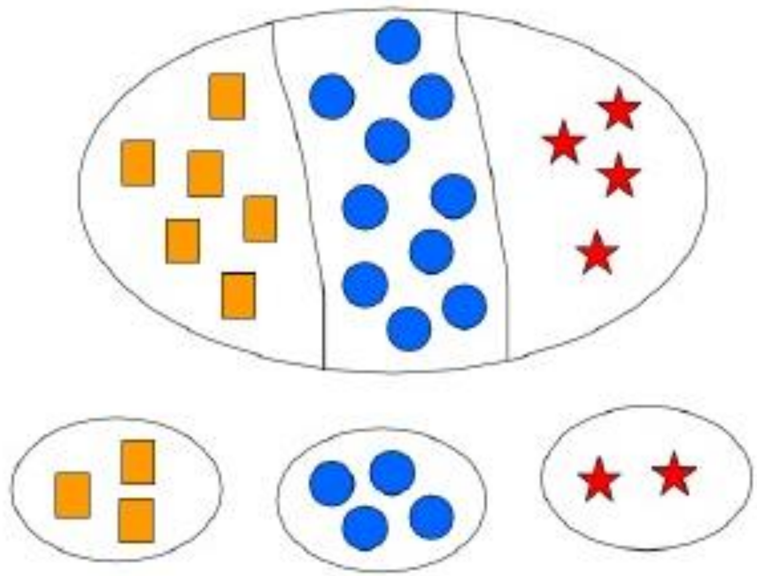
III
Muestreo Aleatorio Estratificado
(MAE)

Muestreo Aleatorio Estratificado (MAE)

Características:

- Este tipo de muestreo se utiliza especialmente cuando se sospecha que la población es heterogénea en cuanto a alguna característica asociada a las variables de estudio.
- Esto obliga a dividir a la población en subpoblaciones o estratos de acuerdo a la variabilidad de esta característica, con el objeto de mejorar las estimaciones.
- Se realiza una MAS de cada una de los estratos.
- Para generar este tipo de muestreo es necesario identificar dentro de la población cada subpoblación o estrato y luego realizar una selección aleatoria simple de los elementos al interior de cada una de estas subpoblaciones.
- Método que permite reducir los costos, es definiendo estrato.

Para conformar los estratos es importantes que los elementos sean homogéneos en su interior, diferentes entre si en propiedades y tamaño.



Estrato 1 Estrato 2 Estrato 3

Nota: Los estratos más grandes tendrán mayor probabilidad de ser representados

Tipos de estratos

1. *Afijación uniforme:* selecciona la misma cantidad de elementos en cada uno de los estratos (Ei)

2. *Afijación óptima:* caso particular del anterior, consiste en seleccionar la muestra de tal manera que los estrato más heterogéneos tengan mas casos. Ello requiere conocer la variabilidad entre estratos.

3. *Afijación de mínima varianza:* se eligen los tamaños de tal manera que se minimice la varianza del estimador, para un coste especificado.

4. *Afijación proporciona:* el tamaño de la muestra de cada estrato es proporcional al tamaño del estrato correspondiente.

Afijación uniforme

- Selecciona la misma cantidad de elementos en cada uno de los estratos (Ei)
- Ajusta convencionalmente los tamaños de los estratos muestrales para aumentar la eficiencia de la selección de los grupos más pequeños.
- Esta condición se deberá tener en cuenta al hacer inferencias (corregir las inferencias).

Muestreo Estratificado no proporcional

Ejemplo, estudio comparativo de acceso de servicios de salud entre personas que viven en municipios pequeños, medianos y grandes.

- Si se emplea un muestreo MAS muy poca gente se encontraran en la muestra pequeña.
- Muestras muy pequeñas => error muestral grande => imposible realizar inferencia y comparación.
- La solución, es aplicar un MAE con muestras del mismo tamaño de personas que vivan en municipios pequeños, medianos y grandes.

1. Afijación proporcional:

- El número de unidades de análisis, seleccionado de cada estrato, es proporcional al número de elementos en cada estrato para la población.

- Establece la distribución proporcional del universo y aplica esta distribución a su tamaño muestral para conformar estratos en la muestra.

- Se eligen aleatoriamente los elementos al interior de cada estrato hasta ajustar su tamaño.

- Es mejor que el MAS pues disminuye el error estándar de la medición muestral.

Afijación óptima

- Caso particular del anterior, consiste en seleccionar la muestra de tal manera que los estrato más heterogéneos tengan mas casos. Ello requiere conocer la variabilidad entre estratos.

- Selecciona el tamaño de los estratos en función de la desviación estándar de cada uno de ellos, de tal manera que los estratos más heterogéneos (mayores varianzas) aporten más casos a la muestra total.

Ejemplo 1. - suponga que se realiza una encuesta donde a una empresa cuenta con tiempo y dinero suficientes para entrevistar n= 40 hogares y decide seleccionar m.a de tamaño n_1 =20 del pueblo A, n_2 = 8 del pueblo B y n_3 = 12 del área rural. Se seleccionan las muestras irrestrictas aleatorias y se realizan las entrevistas.

Los resultados, de mediciones del tiempo en horas por semana que se ve TV se muestran en las siguientes tablas.

E1 Región I	**E2 Región II**	**E3 Región III**
35 28 26 41 43 29 32 37 36 25 29 31 39 38 40 45 28 27 35 34	27 4 49 10 15 41 25 30	8 15 21 7 14 30 20 11 12 32 34 24

E1	**E2**	**E3**
$n_1 = 20$ $\overline{y} = 33.900$ $s_1^2 = 35.358$ $N_1 = 155$	$n_2 = 8$ $\overline{y}_2 = 25.125$ $s_2^2 = 232.411$ $N_2 = 62$	$n_3 = 12$ $\overline{y}_3 = 19.00$ $s_3^2 = 87.636$ $N_3 = 93$

i. Estime el tiempo promedio que se ve televisión, en horas por semana para (a) los hogares de la Región I y (b) hogares de la Región II.
ii. Fije un límite para el error de estimación.
iii. Estime τ.
iv. Fije un límite para el error de estimación.

Solución

A). de los valores de la segunda tabla y usando

$$\overline{y_{st}} = \frac{1}{N}\left[N_1\overline{y_1} + N_2\overline{y_2} + \ldots\ldots + N_L\overline{y_L}\right] = \frac{1}{N}\sum_{i=1}^{L} N_i\overline{y_i}$$

$$= \overline{y_{st}} = \frac{1}{N}\left[N_1\overline{y_1} + N_2\overline{y_2} + N_3\overline{y_3}\right] = \frac{14}{310}\left[(155)(33.900) + (62)(25.125) + (93)(19.000)\right]$$

$$= 27.7$$

- Es la mejor estimación del número promedio de horas por semana en que en todos los hogares de la ciudad ve TV.

$$\hat{V}(\overline{y_{st}}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\left(\frac{N_1 - n_1}{N_1}\right)\left(\frac{s_1^2}{n_1}\right) = \frac{1}{(310)^2}\left[\frac{(155)^2(0.87)(35.358)}{20} + \frac{(62)^2(0.871)(232.411)}{8} + \frac{(93)^2(0.871)(87.636)}{12}\right]$$

$$= 1.97$$

- La estimación de la media poblacional, con un limite para el error de estimación al nivel del 0.95, esta dada por

$$\overline{y_{st}} \pm 2\sqrt{\hat{V}(\overline{y_{st}})} = 27.7 \pm 2\sqrt{1.97} = 27.7 \pm 2.8$$

- Entonces se estima que el número promedio de horas que se ve televisión en los hogares del ciudad= 27.7 hrs.
- Error de estimación= 2.8 hrs con una probabilidad de 0.95.

B). Las n=8 observaciones del E2 provienen de una MAI, por lo tanto la estimación del tiempo promedio de ver TV en la Región II, su error de estimación es:

$$\bar{y}_2 \pm 2\sqrt{\left(\frac{N_2 - n_2}{N_2}\right)\left(\frac{s_2^2}{n_2}\right)} = 25.1 \pm 2\sqrt{\left(\frac{62-8}{62}\right)\left(\frac{232.411}{8}\right)} = 25.1 \pm 10.1$$

- Se observa un límite grande para el error de estimación debido a la presencia de una varianza amplia y un tamaño de “n” pequeño.
- No obstante, la estimación de μ es buena pero la media del estrato no.
- Por lo tanto, si se desea una estimación para un estrato en particular, su “n” debe ser lo bastante grande para proporcionar un límite de error de estimación razonable.

C). Se obtiene:

$$N\overline{y}_{st} = 310(27.7) = 8{,}587 \text{ horas}$$

Con varianza estimada:

$$\hat{V}(N\overline{y}_{st}) = N^2\hat{V}(\overline{y}_{st}) = (310)^2(1.97) = 189{,}278.56$$

D). Por lo que, la estimación del total del número de horas que la población dedica a ver TV, con un limite para el error de estimación será:

$$N\overline{y}_{st} \pm 2\sqrt{\hat{V}(N\overline{y}_{st})} = 8{,}587 \pm 2\sqrt{189{,}278.56} =$$

$$= 435.068x(2) = 8{,}587 \pm 870$$

Ejemplo 2.- , de la encuesta anterior sugiere que las varianzas de los estratos del ejemplo 1 son aproximadamente.

$$\sigma_1^2 \approx 25, \quad \sigma_2^2 \approx 225, \quad \text{y} \ , \sigma_3^2 \approx 100$$

- Estimar μ.
- Obtener un límite en el error de estimación igual a 2 horas

Si las fracciones asignadas son

$$w_1 = \frac{1}{3}, w_2 = \frac{1}{3} \quad \text{y} \quad w_3 = \frac{1}{3}$$

- Nota suponga que toma un igual número de observaciones para cada Ei.

Solución

- ❑ Limite de error de estimación de 2 hrs implica que

$$2\sqrt{\hat{V}(\bar{y}_{st})} = 2$$

$$\Rightarrow \hat{V}(\bar{y}_{st}) = 1$$

- ❑ Por lo tanto D = 1

Se sabe que: $N_1 = 155, \quad N_2 = 62 \quad$ y $N_3 = 93.$

Por lo tanto $$\sum_{i=1}^{3} \frac{N_i^2 \sigma_i^2}{w_i} = \frac{N_1^2 \sigma_1^2}{w_1} + \frac{N_2^2 \sigma_2^2}{w_2} + \frac{N_3^2 \sigma_3^2}{w_3}$$

$$= \frac{(155)^2 (25)}{\frac{1}{3}} + \frac{(62)^2 (225)}{\frac{1}{3}} + \frac{(93)^2 (100)}{\frac{1}{3}} = (24025)(75) + (3844)(675) + (8649)(300)$$

$$= 6991275$$

$$\sum_{i=1}^{3} N_i \, \sigma_i^2 = N_1 \, \sigma_1^2 + N_2 \sigma_2^2 + N_3 \sigma_3^2 = (155)(25) + (62)(225) + (93)(100) = 27125$$

N^2 D = $(310 \quad)^2 (1) = 96100$

Entonces:
$$n = \frac{\sum_{i=1}^{L} N_i^2 \sigma_i^2 \big/ w_i}{N^2 D + \sum_{i=1}^{L} N_i \, \sigma_i^2} = \frac{6991275}{96100 + 27125} = \frac{6991275}{123225} = 56.7$$

Por lo que se debe tomar n= 57 observaciones distribuidas de la siguiente manera:

$$n_1 = n(w_1) = 57\left(\frac{1}{3}\right) = 19$$
$$n_2 = 19$$
$$n = 19$$

- Por otra parte, la empresa tiene el interés de estimar la P de hogares en el ciudad donde se ve el programa X. Recuerde que la Cd. Esta dividida en tres estratos: E1, E2, y E3.
- Los estratos contienen $N1 = 155$, $N_2 = 62$ y $N_3 = 93$. hogares, respectivamente. Una muestra aleatoria estratificada de n = 40 se toma una MIA de cada estrato de la siguiente manera: $n_1 = 20$, $n_2 = 8$ y $n_3 = 12$.
- Las entrevistas son tomadas en los 40 hogares muestreados; los resultados se presentan en la siguiente tabla.

a). Estimar $\hat{p}_i$ de hogares donde se ve el programa X.

b). Fije un límite para el error de estimación.

Ei	**n**	**Núm. de hogares donde se ve el programa X**	**p**
1	$n_1 = 20$	16	0.80
2	$n_2 = 8$	2	0.25
3	$n_3 = 12$	6	0.50

Solución

$$\hat{p}_{st} = \frac{1}{310}\left((155)(0.80)+(62)(0.25)+(93)(0.50)\right) = 0.60$$

Las varianzas para cada Ei:

$$\hat{V}\left(\hat{p}_1\right) = \left(\frac{N \quad - n_i}{N_i}\right)\left(\frac{\hat{p}\hat{q}}{n_1 - 1}\right) = \left(\frac{155-20}{155}\right)\left(\frac{(0.80)(0.20)}{19}\right) = (0.871)(0.008) = 0.007$$

$$\hat{V}\left(\hat{p}_2\right) = \left(\frac{N_2 - n_2}{N_2}\right)\left(\frac{\hat{p}\hat{q}}{n_2 - 1}\right) = \left(\frac{62-8}{62}\right)\left(\frac{(0.25)(0.75)}{7}\right) = (0.871)(0.027) = 0.024$$

$$\hat{V}\left(\hat{p}_3\right) = \left(\frac{N_3 - n_3}{N_3}\right)\left(\frac{\hat{p}\hat{q}}{n_3 - 1}\right) = \left(\frac{93-12}{93}\right)\left(\frac{(0.5)(0.50)}{11}\right) = (0.871)(0.023) = 0.020$$

Entonces la varianza es

$$\hat{V}\left(\hat{p}_{st}\right) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \hat{V}\left(\hat{p}_i\right) = \frac{1}{(310)^2}\left[(155)^2(0.007)+(62)^2(0.024)+(93)^2(0.020)\right] = 0.0045$$

Error de estimación es $2\sqrt{\hat{V}\left(\hat{p}_{st}\right)} = 2(0.07) = 1.4$

Límite del error de estimación: 0.60 ± 1.4

IV. Muestreo Aleatorio por Conglomerados (MAC)

Características

- Suele denominarse en múltiples etapas
- Suele aplicarse cuando las poblaciones son muy grandes y dispersas
- No se dispone de un marco muestral (listado)
- En lugar de individuos se seleccionan conglomerados que están agrupados de forma natural (manzanas, departamentos, municipios, hospitales, provincias, escuelas etc.)
- Se selecciona en primer lugar el conglomerado mas alto, a partir de ese se selecciona un subgrupo, a partir de ese subgrupo se selecciona otro subgrupo y así sucesivamente hasta llegar a las unidades de análisis.

- Una muestra por conglomerados es una muestra aleatoria en la cual cada unidad de muestreo es una colección, o conglomerado, de elementos.

- El muestreo por conglomerados es un diseño efectivo para obtener una cantidad especificada de información al costo mínimo bajo las siguientes condiciones:

- No se encuentra disponible o es muy costoso obtener un buen marco que liste los elementos de la población, mientras que se puede lograr fácilmente un marco que liste los conglomerados.

- El costo por obtener observaciones se incrementa con la distancia que separa los elementos.

Proceso

1. Se selecciona una muestra de “m” conglomerados mediante MAS.
2. En cada “m” seleccionado se obtiene un marco de las Ni unidades i=1,2,...,m
3. Seleccionar una m.a. de tamaño ni, i=1,2,...,m de cada uno de los “m” conglomerados.

Tamaño de la muestra n=n1+n2+...nm. Para determinar “ni” para cada conglomerado se puede hacer por separado o bien determinar “n” y después distribuirla sobre los “m” conglomerados.

❑ Selección de una muestra

- Primero debe especificar los conglomerados apropiados. Los elementos dentro de un conglomerado están frecuentemente juntos físicamente, por lo que tienden a presentar características similares.

- Cuando los elementos de un conglomerado son muy diferentes entre sí => una muestra que contenga pocos conglomerados grandes puede producir una estimación muy buena de un parámetro poblacional, tal como la media.

- La principal diferencia entre IMAC y MAE es que los estratos deben ser tan homogéneos (semejantes) entre ellos, como sea posible, pero un estrato debe diferir tanto como sea posible de otro con respecto a la característica que está siendo medida.
- Los conglomerados, por otro lado, deben ser tan heterogéneos (diferentes) entre ellos como sea posible, y un conglomerado debe ser muy similar a otro para poder aprovechar las ventajas económicas del muestreo por conglomerados.

- Una vez que los conglomerados han sido especificados se debe conformar un marco que liste todos los conglomerados de la población, después se selecciona una muestra irrestricta aleatoria de conglomerados de este marco.

❑ Muestreo bietapico

Caso 1: aplicar M.A.S. en ambas etapas.

- 1ª etapa: se eligen **n** UPM de las **N** con M.A.S.
- 2ª etapa: se eligen m_i USM con M.A.S de la i-ésima UPM seleccionada.

El total de la j-ésima UPM seleccionada:

$$\hat{y}_j = M_j \hat{\bar{y}}_j \ siendo\ \hat{y}_j = \sum_{i=1}^{m_j} y_{ij}$$

Media poblacional es insesgadamente estimada:

$$\boxed{\hat{\mathbf{y}} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{j=1}}^{\mathbf{n}} \hat{\mathbf{y}}_{\mathbf{j}}}$$

Caso 2: UPM elegidas con proporción al tamaño y USM con M.A.S

El estimador insesgado del total:

$$\boxed{\hat{\mathbf{Y}}_{\mathbf{ppt}} = \frac{\mathbf{M}}{\mathbf{n}} \sum_{\mathbf{j=1}}^{\mathbf{n}} \hat{\mathbf{y}}_{\mathbf{j}}}$$

Varianza estimada:

$$\boxed{\hat{\text{var}}(\hat{Y}_{ppt}) = \frac{M^2}{n(n-1)} \sum_{j=1}^{n} \left(\hat{y}_j - \frac{\hat{Y}_{ppt}}{M} \right)^2}$$

Ejemplos de UPM y de USM

Variable de interés	Unidades Primarias UPM	Unidades secundarias USM
Gasto, ingreso, periódicos o revistas	Manzanas	Familias
Producción de maíz. Carne, leche, etc.	Municipios	Predios
Contenido de calcio , nitrógeno, etc.	Arboles	Hojas o frutos
Trabajo, ingreso, drogadicción	Manzanas	Individuos dentro de la manzana
Calificaciones en la faculta de ciencias	Grupos	Alumnos

❑ Estimación de la media μ y el Total poblaciona τ

El muestreo por conglomerados es muestreo irrestricto aleatorio con cada unidad de muestreo conteniendo un número de elementos. Los estimadores de la media poblacional μ y el total τ son similares a los de muestreo irrestricto aleatorio. La media muestral $\bar{y}$ es un buen estimador de la media poblacional μ. En esta sección se estudia un estimador de μ y dos estimadores de τ.

N = número de *conglomerados* en la población.

$n =$ Número de conglomerados seleccionados en una muestra irrestricta aleatoria.

$m_i =$ Número de elementos en el conglomerado $i, i, = 1, \ldots, N$

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i = \text{tamaño promedio del conglomerado en la muestra.}$$

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

$$\bar{M} = \frac{M}{N} = \text{tamaño promedio del conglomerado en la población.}$$

$y_i =$ total de todas las observaciones en el $i -$ésimo conglomerado.

El estimador de la media poblacional μ es la media muestral $\bar{y}$, la cual es dada por:

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

La media $\bar{y}$ toma la forma de un estimador de razón, con m_i. La varianza estimada de $\bar{y}$ toma la forma de la varianza de un estimador de razón.

- **Estimador de la media poblacional μ :**

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

- **Varianza estimada de $\bar{y}$:**

$$\hat{V}(\bar{y}) = \left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\,m_i\right)^2}{n-1}$$

- **Límite para el error de estimación**

$$2\sqrt{\hat{V}\left(\bar{y}\right)} = 2\sqrt{\left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\,m_i\right)^2}{n-1}}$$

Aquí $\bar{M}$ puede ser estimado por $\bar{m}$ si se desconoce M.

La varianza estimada es sesgada y sería un buen estimador de $V(\bar{y})$ únicamente si n fuera grande, digamos $n \geq 20$. El sesgo desaparece cuando los tamaños de los conglomerados $m_1, m_2, \ldots, m_N$ son iguales.

- **Estimador del total poblacional τ :**

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

- **Varianza estimada de $M\,\bar{y}$:**

$$\hat{V}(M\,\bar{y}) = M^2\hat{V}(\bar{y}) = N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\,m_i\right)^2}{n-1}$$

- **Límite para el error de la estimación:**

$$2\sqrt{\hat{V}(M\,\bar{y})} = 2\sqrt{N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\,m_i\right)^2}{n-1}}$$

- **Estimador del total poblacional τ, el cual no depende de M :**

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

- **Varianza estimada de $N\bar{y}$:**

$$\hat{V}(N\bar{y}_t) = N^2\hat{V}(\bar{y}_t) = N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_t\right)^2}{n-1}$$

- **Límite para el error de estimación:**

$$2\sqrt{\hat{V}(N\bar{y}_\tau)} = 2\sqrt{N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_t\right)^2}{n-1}}$$

Si existe una gran cantidad de variación entre los tamaños de los conglomerados y si los tamaños están altamente correlacionados con los totales de conglomerados, la varianza del total poblacional es generalmente mayor que la varianza de total del conglomerad . El estimador no usa la información proporcionada por los tamaños de los conglomerados y por esto puede ser menos preciso.

□ SELECCIÓN DEL TAMAÑO DE MUESTRA PARA μ Y τ

La cantidad de información en una muestra por conglomerados es afectada por dos factores, el número y el tamaño relativo de los conglomerados.

□ Varianza estimada de $\bar{y}$:

$$\hat{V}(\bar{y}) = \frac{N-n}{Nn\bar{M}^2}(s_c^2)$$

Donde $s_c^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\,m_i\right)^2}{n-1}$

□ La varianza real de $\bar{y}$ es aproximadamente

$$V(\bar{y}) = \frac{N-n}{Nn\bar{M}^2}(\sigma_c^2)$$

Donde σ^2 es la cantidad poblacional estimada por s_c^2 .

Debido a que no conocemos σ^2 o el tamaño promedio $\bar{M}$ del conglomerado, la elección del tamaño de muestra, esto es, el número de conglomerados necesario para comprar una cantidad especificada de información concerniente a un parámetro poblacional, es complicado.

En los problemas de selección de un tamaño de muestra, igualamos dos desviaciones estándar de nuestro estimador, con un límite para el error B .

$$2\sqrt{V(\bar{y})} = B \quad 2$$

- **Tamaño de muestra aproximado requerido para estimar μ con un límite B para el error de estimación:**

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}$$

Donde σ_c^2 es estimado por s_c^2 y

$$D = \frac{B^2 \bar{M}^2}{4}$$

- **Tamaño de muestra aproximado requerido para estimar τ, usando $M\bar{y}$, con un límite B para el error de estimación:**

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}$$

Donde σ_c^2 es estimada por s_c^2 y

Donde σ_c^2 es estimada por s_c^2 y

$$D = \frac{B^2}{4N^2}$$

El estimador $N\bar{y}_t$ que se muestra en la ecuación, se usa para estimar τ cuando M es desconocido. La varianza estimada de $N\bar{y}_t$ que se muestra en la ecuación, es

$$\hat{V}(N\bar{y}_t) = N^2\left(\frac{N-n}{Nn}\right)s_t^2$$

Donde $$s_t^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_t\right)^2}{n-1}$$

Entonces la varianza poblacional de $N\bar{y}_t$ es

$$V(N\bar{y}_t) = N^2V(\bar{y}_t) = N^2\left(\frac{N-n}{Nn}\right)\sigma_t^2$$

- **Tamaño de muestra aproximado requerido para estimar τ, usando $N y_t$ con un límite B para el error de estimación:**

$$n = \frac{N\sigma_t^2}{ND + \sigma_t^2}$$

Donde σ_t^2 se estima mediante s_t^2, y

$$D = \frac{B^2}{4N^2}$$

❑ ESTIMACIÓN DE UNA PROPORCIÓN (p) POBLACIONAL

El mejor estimador de la proporción poblacional “p” es la proporción muestral. Sea “a_i” el número total de elementos en el conglomerado “i” que poseen la característica de interés. Entonces la proporción de elementos en la muestra “n” conglomerados que poseen la característica de interés dada por:

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

Donde mi, es eel numero de elementos en el i-ésimo conglomerado, i=1,2, … n

- **Estimador de la proporción poblacional p :**

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

- **Varianza estimada de $\hat{p}$:**

$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^{n} (a_i - \hat{p} m_i)^2}{n-1}$$

- Límite para el error de estimación:

$$2\sqrt{\hat{V}(\hat{p})} = 2\sqrt{\left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^{n} (a_i - \hat{p} m_i)^2}{n-1}}$$

❑ TAMAÑO DE MUESTRA PARA (P)

La estimación de la proporción poblacional , con un límite de unidades para el error de estimación, implica que el experimentador quiere

$$2\sqrt{V(\hat{p})} = B$$

Esta ecuación puede ser resuelta para :

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}$$

Donde D=B^2 , y se estima por:

$$s_c^2 = \frac{\sum_{i=1}^{n}(a_i - pm_i)^2}{n-1}$$

M.A. C. COMBINADO CON ESTRATIFICACIÓN

- El muestreo por conglomerados puede ser combinado con muestreo estratificado, con objeto de que la población pueda ser dividida en L estratos y se pueda seleccionar una muestra por conglomerados en cada estrato.

- Planteado en términos de un estimador de razón, existen dos formas para construir el estimador de una media poblacional a través de los estratos: el estimador separado y el estimador combinado.

El estimador del promedio poblacional del total por conglomerado es:

$$\frac{1}{N}\left(N_1\,\bar{y}_{t1}+N_2\,\bar{y}_{t2}\right)$$

Mientras que el estimador del promedio del tamaño de conglomerados es:

$$\frac{1}{N}\left(N_1\,\bar{m}_1 + N_2\,\bar{m}_2 \right)$$

Un estimador de la media poblacional por elementos es:

$$\bar{y}^* = \frac{N_1\,\bar{y}_{t1} + N_2\,\bar{y}_{t2}}{N_1\,\bar{m}_1 + N_2\,\bar{m}_2}$$

y esta ecuación tiene la forma de un estimador de razón combinada. La varianza de puede ser estimada por:

$$\hat{V}(\bar{y}^*) = \frac{1}{M^2}\left\{ \frac{N_1(N-n_1)}{n_1(n_1-1)} \sum_{i=1}^{n_1}\left[\left(y_i - \bar{y}_{t1}\right) - \bar{y}^*\left(m_i - \bar{m}_1\right)\right]^2 + \frac{N_2(N_2-n_2)}{n_2(n_2-1)} \sum_{i=1}^{n_2}\left[\left(y_i - \bar{y}_{t2}\right) - \bar{y}^*\left(m_i - \bar{m}_2\right)\right]^2 \right\}$$

Donde: M es el número total de elementos en la población y puede ser estimado por $N_1\,\bar{m}_1 + N_2\,\bar{m}_2$ si no es conocido. La primera suma en la expresión de la varianza es sobre todas las observaciones de la muestra en el estrato 1, y la segunda suma es sobre todas las observaciones del estrato 2.

MAC CON PROBABILIDADES PROPORCIONALES (ppt) AL TAMAÑO

El MAC suele proporcionar una situación ideal para el uso de muestreo ppt, ya que el número de elementos en un conglomerado “m_i” , representa una medida natural del tamaño del conglomerado.

El muestreo con probabilidades proporcionales a “m_i” devenga beneficios en términos de la reducción del límite para el error de estimación, cuando el total del conglomerado “y_i” está altamente correlacionado con el número de elementos en el conglomerado, lo cual ocurre frecuentemente.

Sea π_i , la probabilidad de que la i-ésima unidad de muestreo aparezca en la muestra, la cual es dada por: $\pi_i = \dfrac{m_i}{M}$

Entonces, el estimador de un total poblacional $\hat{\tau}_{ppt}$ toma la forma:

$$\hat{\tau}_{ppt} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\pi_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{(m_i / M)}$$

$$= \frac{M}{n}\sum_{i=1}^{n}\frac{y_i}{m_i} = \frac{M}{n}\sum_{i=1}^{n}\bar{y}_i$$

$\bar{y}_i$

Donde $\bar{y}_i$ Escriba aquí la ecuación.es el promedio de las observaciones en el i-ésimo conglomerado. La varianza estimada de $\hat{\tau}_{ppt}$ tiene una forma particularmente simple.

Ya que ahora hay M elementos en la población, el estimador de la media poblacional, $\hat{\mu}_{ppt}$, es:

$$\hat{\mu}_{ppt} = \frac{1}{M}\hat{\tau}_{ppt} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$$

- **Estimador de la media poblacional μ :**

$$\hat{\mu}_{ppt} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$$

donde $\bar{y}_i$ es la media del i-ésimo conglomerado.

- **Varianza estimada de $\hat{\mu}_{ppt}$:**

$$\hat{V}(\hat{\mu}_{ppt}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\bar{y}_i - \hat{\mu}_{ppt}\right)^2$$

- **Límite para el error de estimación:**

$$2\sqrt{\hat{V}(\hat{\mu}_{ppt})} = 2\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\bar{y}_i - \hat{\mu}_{ppt}\right)^2}$$

- **Estimador del total poblacional τ :**

$$\hat{\tau}_{ppt} = \frac{M}{n}\sum_{i=1}^{n}\bar{y}_i$$

- **Varianza estimada de $\hat{\tau}_{ppt}$:**

$$\hat{V}(\hat{\tau}_{ppt}) = \frac{M^2}{n(n-1)}\sum_{i=1}^{n}\left(\bar{y}_i - \hat{\mu}_{ppt}\right)^2$$

- **Límite para el error de estimación:**

$$2\sqrt{\hat{V}(\hat{\tau}_{ppt})} = 2\sqrt{\frac{M^2}{n((n-1)}\sum_{i=1}^{n}\left(\bar{y}_i - \mu_{ppt}\right)^2}$$

Ejemplo 1. Con los siguientes datos de ingresos se pide:

a) Estimar el ingreso promedio por persona en la ciudad y establezca un límite para el error de estimación.

b) Si existen 2,500 residentes en la ciudad. Estime el ingreso total de todos los residentes de la ciudad y ponga un límite para el error de estimación.

Conglomerado	Número de residentes	Ingreso total por conglomerado	Conglomerado	Número de residentes	Ingreso total por conglomerado	Número de arrendatarios	Número de arrendatarios
1	8	$ 96,000	14	10	$ 49,000	4	5
2	12	121,000	15	9	53,000	7	4
3	4	42,000	16	3	50,000	1	1
4	5	65,000	17	6	32,000	3	4
5	6	52,000	18	5	22,000	3	2
6	6	40,000	19	5	45,000	4	3
7	7	75,000	20	4	37,000	4	1
8	5	65,000	21	6	51,000	2	3
9	8	45,000	22	8	30,000	3	3
10	3	50,000	23	7	39,000	2	4
11	2	85,000	24	3	47,000	1	0
12	6	43,000	25	8	41,000	3	3
13	5	54,000				2	

Solución

a).Estimar el ingreso promedio por persona en la ciudad y establezca un límite para el error de estimación.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} = \frac{\$1{,}329{,}000}{151} = \$8{,}801$$

Calculando $\hat{V}(\bar{y})$

$$\sum_{i=1}^{25} y_i^2 = y_1^2 + y_2^2 + \ldots + y_{25}^2$$

$$= (96{,}000)^2 + (121{,}000)^2 + \ldots + (41{,}000)^2$$

$$= 82{,}039{,}000{,}000$$

$$\sum_{i=1}^{25} m_i^2 = m_1^2 + m_2^2 + \ldots + m_{25}^2$$

$$= (8)^2 + (12)^2 + \ldots + (8)^2 = 1{,}047$$

$$\sum_{i=1}^{25} y_i m_i = y_1 m_1 + y_2 m_2 + \ldots + y_{25} m_{25}$$

$$= (96{,}000)(8) + (121{,}000)(12) + \ldots + (41{,}000)(8)$$

$$= 8{,}403{,}000$$

Igualando:

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

Sustituyendo:

$$\sum_{i=1}^{25}\left(y_i - \bar{y}m_i\right)^2 = 82{,}039{,}000{,}000 - 2(8801)(8{,}403{,}000) + (8801)^2(1047) \qquad = 15{,}227{,}502{,}247$$

M se desconoce, la $\bar{M}$ debe ser estimada por $\bar{m}$:

$$\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n} = \frac{151}{25} = 6.04 \qquad N = 415$$

$$\hat{V}(\bar{y}) = \left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}m_i\right)^2}{n-1} \qquad = \left[\frac{415-25}{(415)(25)(6.04)^2}\right]\left(\frac{15{,}227{,}502{,}247}{24}\right) = 653{,}785$$

La estimación de μ con un límite para el error de estimación es: $\bar{y} \pm 2\sqrt{\hat{V}(\bar{y})}$
o sea $8801 \pm \sqrt{653{,}785}$ $\quad 8801 \pm 1617$

La estimación del ingreso promedio por persona es $8,801, y el error de estimación debe ser menor que $1617 con una probabilidad cercana a 0.95.

b) Si existen 2,500 residentes en la ciudad. Estime el ingreso total de todos los residentes de la ciudad y ponga un límite para el error de estimación.

La media muestral $\bar{y}$ se calcula de \$ 8801 (A). La estimación de τ es:

$$M\bar{y} = 2500(8801) = \$22{,}002{,}500$$

La estimación de τ con un límite para el error de estimación es:

$$M\bar{y} \pm 2\sqrt{\hat{V}(M\bar{y})} = M\bar{y} \pm 2\sqrt{M^2\hat{V}(\bar{y})}$$

$$22{,}002{,}500 \pm 2\sqrt{(2500)^2(653{,}785)}$$

$$22{,}002{,}500 \pm 4{,}042{,}848$$

B).Estimar el ingreso total de todos los residentes de la ciudad si M no es conocido. Establezca un límite para el error de estimación.

$N = 415$, la estimación del ingreso total τ es:

$$\text{¡} N\bar{y}_t = \frac{N}{n}\sum_{i=1}^{n} y_i = \frac{415}{25}(1{,}329{,}000) = \$22{,}061{,}400$$

Límite al error de estimación, primero calculamos:

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

$$= 82{,}039{,}000{,}000 - \frac{1}{25}(1{,}329{,}000)^2$$

$$= 11{,}389{,}360{,}000$$

La estimación del ingreso total de todos los residentes de la ciudad, con un límite para el error de estimación, es:

$$N\bar{y}_t \pm 2\sqrt{\hat{V}(N\bar{y}_t)}$$

Sustituyendo en la ecuación

$$N\bar{y}_t \pm 2\sqrt{N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_t\right)^2}{n-1}}$$

$$22{,}061{,}400 \pm 2\sqrt{(415)^2\left[\frac{415-25}{(415)(25)}\right]\left(\frac{11{,}389{,}360{,}000}{24}\right)}$$

$$22{,}061{,}400 \pm 3{,}505{,}920$$

¿Qué tan grande debe tomarse la muestra en una encuesta futura para estimar el ingreso promedio por persona μ con un límite de \$ 500 para el error de estimación?

$$s_c^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}m_i\right)^2}{n-1} = \frac{15{,}227{,}502{,}247}{24} = 634{,}479{,}260$$

$\bar{M}$ puede ser estimada por $\bar{m} = 6.04$. Entonces D es aproximadamente

$$\frac{B^2\bar{m}^2}{4} = \frac{(500)^2(6.04)^2}{4} = (62{,}500)(6.04)^2$$

Usando la ecuación (8.13):

$$n = \frac{N\sigma_c^2}{ND+\sigma_c^2} = \frac{415(634{,}479{,}260)}{415(6.04)^2(62{,}500) + 634{,}479{,}260} = 166.58$$

Entonces se deben muestrear 167 conglomerados.

¿Qué tan grande se necesita una muestra para estimar el ingreso total de todos los residentes, τ , con un límite de \$1,000,000 para el error de estimación? $(M = 2500)$

Usando la ecuación para estimar σ_c^2 :

$s_c^2 = 634{,}479{,}260$

Cuando estimamos τ , usamos:

$$D = \frac{B^2}{4N^2} = \frac{(1{,}000{,}000)^2}{4(415)^2}$$

$$ND = \frac{(1{,}000{,}000)^2}{4(415)} = 602{,}409{,}000$$

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2} = \frac{415(634{,}479{,}260)}{602{,}409{,}000 + 634{,}479{,}260} = 212.88$$

Se debe muestrear 213 conglomerados para estimar el ingreso total con un límite de \$1,000, 000 para el error de estimación.

Entonces se deben muestrear 167 conglomerados.

¿Qué tan grande se necesita una muestra para estimar el ingreso total de todos los residentes, τ , con un límite de $1,000,000 para el error de estimación? $(M = 2500)$

Usando la ecuación para estimar σ_c^2 :

$s_c^2 = 634{,}479{,}260$

Cuando estimamos τ , usamos:

$$D = \frac{B^2}{4N^2} = \frac{(1{,}000{,}000)^2}{4(415)^2}$$

$$ND = \frac{(1{,}000{,}000)^2}{4(415)} = 602{,}409{,}000$$

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2} = \frac{415(634{,}479{,}260)}{602{,}409{,}000 + 634{,}479{,}260} = 212.88$$

Se debe muestrear 213 conglomerados para estimar el ingreso total con un límite de $1,000, 000 para el error de estimación.

El mejor estimador de la proporción poblacional de arrendatarios es $\hat{p}$, ecuación (8.18), donde:

$$\hat{p} = \frac{\sum_{i=1}^{m} a_i}{\sum_{i=1}^{n} m_i} = \frac{72}{151} = 0.48$$

Para estimar la varianza de $\hat{p}$, debemos calcular

$$\sum_{i=1}^{n} (a_i - \hat{p}m_i)^2 = \sum_{i=1}^{n} a_i^2 - 2\hat{p}\sum_{i=1}^{n} a_i m_i + \hat{p}^2 \sum_{i=1}^{n} m_i^2$$

$$\sum_{i=1}^{n} (a_i - \hat{p}m_i)^2 = 262 - 2(0.477)(511) + (0.477)^2 (1047) = 12.729$$

$\bar{M}$ es estimada por $\bar{m}$, donde

$$\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n} = \frac{151}{25} = 6.04$$

De

$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum (a_i - \hat{p}m_i)^2}{n-1}$$

$$= \frac{(415-25)(12.729)}{415(25)(6.04)^2(24)} = 0.00055$$

La estimación de p con un límite para el error de estimación es:

$\hat{p} \pm 2\sqrt{\hat{V}(\hat{p})}$, o sea $0.48 \pm 2\sqrt{0.00055}$, o sea 0.48 ± 0.05

La mejor estimación de la proporción de personas que alquilan casa es 0.48.

El error de estimación debe ser menor que 0.05 con probabilidad de aproximadamente 0.95.

¿Qué tan grande se debe tomar la muestra para estimar p, con un límite de 0.04 en el error de estimación?

Estimador:

$$s_c^2 = \frac{\sum_{i=1}^{n}(a_i - \hat{p}m_i)^2}{n-1} = \frac{12.729}{24} = 0.530$$

$\bar{M}$ Es estimada por $\bar{m} = 6.04$. También D es aproximada por

$$\frac{B^2 \bar{m}^2}{4} = \frac{(0.04)^2(6.04)^2}{4} = 0.0146$$

Entonces
$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2} = \frac{(415)(0.530)}{(415)(0.0146) + 0.530} = 33.40$$

Se deben muestrear 34 conglomerados para estimar p, con un límite de 0.04 para el error de estimación.

Consideremos la muestra del estrato 1, con $N_1 = 415$ y $n_1 = 25$. Se toma una ciudad vecina más pequeña como el estrato 2. Para el estrato 2, $n_2 = 10$ bloques se van a muestrear de $N_2 = 168$. Estime el ingreso promedio por persona en las dos ciudades combinadas, y establezca un límite para el error de estimación, dados los siguientes datos adicionales:

Conglomerado	Número de residentes	Ingreso total por conglomerado
1	2	$ 18,000
2	5	52,000
3	7	68,000
4	4	36,000
5	3	45,000
6	8	96,000
7	6	64,000
8	10	115,000
9	3	41,000
10	1	12,000

El promedio de los totales de los conglomerados en las respectivas muestras son $\bar{y}_{t1} = 53{,}160$ y $y_{t2} = 54{,}700$. El promedio de los tamaños de los conglomerados en las respectivas muestras es $\bar{m}_1 = 6.04$ y $\bar{m}_2 = 4.90$.

$$\bar{y}^* = \frac{415(53{,}160) + 168(54{,}700)}{415(6.04) + 168(4.90)} = 9385$$

Para el estrato 1

$$\left(\frac{1}{n_1 - 1}\right)\sum_{i=1}^{n_1}\left[\left(y_i - \bar{y}_{t1}\right) - \bar{y}^*\left(m_i - \bar{m}_1\right)\right]^2 = 675{,}930{,}246$$

Ya que $N_1\,\bar{m}_1 + N_2\,\bar{m}_2 = 3329.8$

Por lo que $\hat{V}(\bar{y}^*) = 412{,}563.8$

Y $2\sqrt{\hat{V}\left(\bar{y}^*\right)} = 1285$

El ingreso promedio por persona para las dos ciudades combinadas es: $9,385 ± $1,285

Bibliografía

1. Cochran, William. (1985) "Técnicas de Muestreo". Compañía Editorial Continental, S.A. México.

2. Downie, M. (1973) "Métodos Estadísticos Aplicados". Harper & Row Publishers INC. México.

3. Lewis, Alvin. (2000) "Bioestadística". Compañía Editorial Continental, S.A. México. S/F.

4. Neter y Otros.(2011), "Fundamentos de Estadística para Economía y Negocios". Compañía Editorial Continental, S.A. México. S/F.

5. Martín Pliego, F. y Ruiz-Maya, L. (1995) *Estadística II: Inferencia*. Madrid: AC.

6. Méndez, I., & Quintana, C. R. H. (2007). Muestreo: Respuesta aleatorizada.

En:http://www.dpye.iimas.unam.mx/finales2007/MuestreoRespuestaAleatoriada. Especialidad en Estadística Aplicada. IIMAS, UNAM.

6. Méndez, I., Eslava, G., & Romero, P. (2004). Conceptos Básicos de Muestreo. Cd México: IIMAS, UNAM.

7. Levy, P. S., Lemeshow, S. (1991), Sampling of Populations Methods and Applications. John Wiley & Sons.

8. Pérez, L. C. (2000). Técnicas de muestreo estadístico. Teoría, práctica y aplicaciones informáticas. México, DF: Alfaomega-Rama.

9. Rendón, S. G. (1997). Métodos estadísticos. Muestreo, diseños experimentales, estadística no paramétrica. México, DF: Universidad Autónoma de Chapingo.

10. Scheaffer, R. L., Mendenhall, W., & Lyman, O. (1987). Elementos de muestreo. México, DF: Grupo Editorial Iberoamérica.

11. Stevenson, William. (1981) "Estadística para Administración y Economía". Harla. México.