TÉCNICAS DE MUESTREO I

MUESTREO ESTRATIFICADO

Profesor: Ing. Celso Gonzales Ch. Mg.Sc

CONTENIDO

- Introducción
- Muestreo Estratificado. Definiciones. Estimación de parámetros, tipos de asignación.

CARACTERISTICAS

- La población se divide en grupos homogéneos (estratos).
- Se selecciona una muestra aleatoria de cada estrato
- Permite utilizar información a priori

Existen tres razones importantes para utilizar este tipo de muestreo:

- estadísticas,
- marcos;
- Costos.

RAZÓN ESTADÍSTICA PARA USAR ESTRATOS:

- Aportar información más precisa de algunas subpoblaciones.
- Uso adecuado, puede generar ganancias en precisión
- Requerimiento de estimaciones para ciertas áreas Existencia de una variable precisa para la estratificación.

MUESTREO ESTRATIFICADO

La población de N unidades se divide en L subpoblaciones (o estratos), en función de cierta característica relevante, de tamaños N_1, N_2,N_3, ...,N_l.

Estratos(h)	Elementos	N_h	W_h	$\bar{Y}_h$	S_h^2
1	$Y_{11},\ldots\ldots Y_{1N_1}$	N_1	W_1	$\bar{Y}_1$	S_1^2
2	$Y_{21},\ldots\ldots Y_{2N_2}$	N_2	W_2	$\bar{Y}_2$	S_2^2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
L	$Y_{L1},\ldots\ldots Y_{LN_L}$	N_L	W_L	$\bar{Y}_L$	S_L^2

Estratos(h)	**Muestra aleatoria**	$\mathbf{n_h}$	$\mathbf{w_h}$	$\bar{y}_h$	s_h^2	$v(\bar{y}_h)$
1	$y_{11},\ldots\ldots y_{1n_1}$	$\mathbf{n_1}$	$\mathbf{w_1}$	$\bar{y}_1$	s_1^2	$v(\bar{y}_1)$
2	$y_{21},\ldots\ldots y_{2n_2}$	$\mathbf{n_2}$	$\mathbf{w_2}$	$\bar{y}_2$	s_2^2	$v(\bar{y}_2)$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
L	$y_{L1},\ldots\ldots y_{Ln_L}$	$\mathbf{n_L}$	$\mathbf{w_L}$	$\bar{y}_L$	s_L^2	$v(\bar{y}_L)$

VENTAJAS:

- **Asegura que la muestra represente adecuadamente la población en función de los criterios de estratificación.**
- **Se obtienen estimaciones más precisas que el muestreo aleatorio simple.**

DESVENTAJAS

- **Se de ha conocer la distribución en la población de las variables utilizadas para la estratificación.**

PROPIEDADES DE LA ESTIMACIONES

- TEOREMA 1

$$E(\overline{y}_{st}) = \overline{Y}$$

- TEOREMA 2 $$V(\overline{y}_{st}) = \sum_{h=1}^{L} W_h^2 V\left(\overline{y}_h\right)$$

ESTIMACION DE PARAMETROS

- **MEDIA**

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

- **VARIANZA**

$$v(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 v(\bar{y}_h)$$

Siendo:

$$v(\bar{y}_h) = (1 - f_h)\frac{\hat{S}_h^2}{n_h}$$

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

- **ERROR ESTANDAR**

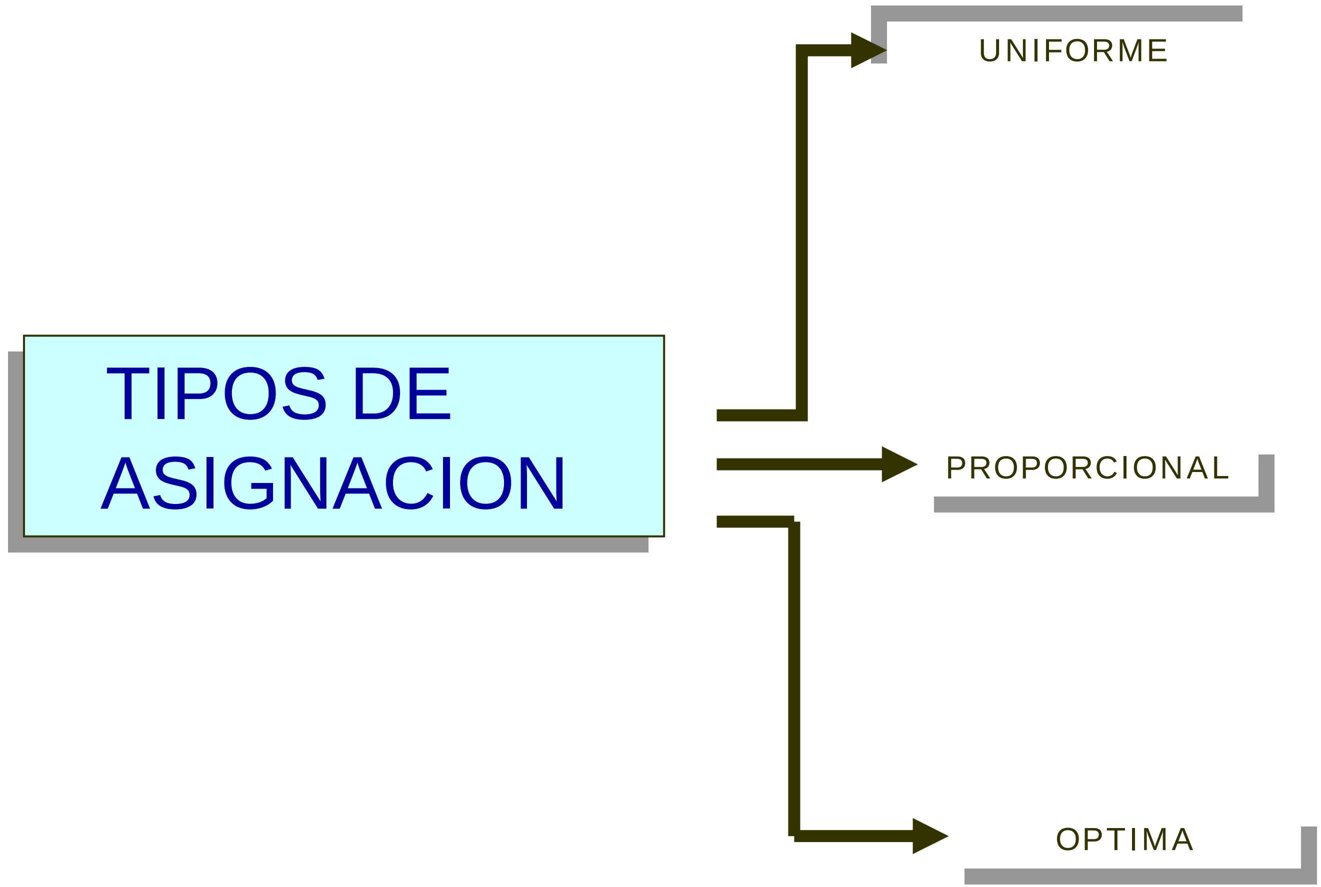
$$\hat{S}_{\bar{y}_{st}} = \sqrt{v(\bar{y}_{st})}$$

Por el *Teorema Límite Central*, para cada estrato, se tendrá que

$$\bar{y}_h \sim N[\bar{Y}_h,\ V(\bar{y}_h)] \qquad \hat{Y} \sim \mathrm{N}\ [Y, V(\hat{Y})]$$

$$LC(\bar{Y}_{st}) = \bar{y}_{st} \pm t s_{\bar{y}_{st}}$$

$$LC(Y) = N\bar{y}_{st} \pm tNs_{\bar{y}_{st}}$$


TIPOS DE ASIGNACION
UNIFORME
PROPORCIONAL
OPTIMA

Asignación Proporcional

$$W_h = w_h$$

$$n_h = n\frac{N_h}{N}$$

Asignación óptima

Minimizar $v(\overline{y}_{st})$

Minimiza la varianza del *estimador*, para un costo especificado, o, habiendo fijado la varianza, minimiza el costo.

$$Min(\phi) = v\left(\overline{y}_{st}\right) + \lambda\left(\sum_{h=1}^{L} n_h - n\right)$$

$$n_h = n\frac{W_h\hat{S}_h}{\sum_{h=1}^{L} W_h\hat{S}_h}$$

ESTIMACION DEL TAMAÑO DE LA MUESTRA DATOS CONTINUOS

Para cualquier asignación: V =(d/t)²

Estimación de la media de población $\bar{Y}$

$$v\left(\bar{y}_{st}\right) = \frac{1}{n}\sum_{h=1}^{L}\frac{W_h^2 \hat{S}_h^2}{w_h} - \frac{1}{N}\sum_{h=1}^{L} W_h \hat{S}_h^2$$

$$n = \frac{\displaystyle\sum_{h=1}^{L}\frac{W_h^2 \hat{S}_h^2}{w_h}}{v\left(\bar{y}_{st}\right) + \dfrac{1}{N}\displaystyle\sum_{h=1}^{L} W_h \hat{S}_h^2}$$

FORMULA GENERAL

Casos particulares:

1. Asignación Proporcional

$$n = \frac{\sum_{h=1}^{L} W_h \hat{S}_h^2}{v\left(\overline{y}_{st}\right) + \frac{1}{N^2}\sum_{h=1}^{L} N_h \hat{S}_h^2}$$

2. Asignación Optima. (n fijo)

$$w_h \ \alpha \ W_h \hat{S}_h$$

$$n = \frac{\left(\sum_{h=1}^{L} W_h \hat{S}_h\right)^2}{v\left(\overline{y}_{st}\right) + \frac{1}{N}\sum_{h=1}^{L} W_h \hat{S}_h^2}$$

EJEMPLO

Suponga que 5000 usuarios de energía eléctrica están divididas de la siguiente manera:

Nivel socioeconomico	Nº de usuarios
A	2500
B	1500
C	1000
Total	5000

Se toma una muestra aleatoria de 50 usuarios seleccionadas de los tres estratos. Usando la asignación proporcional cual es el tamaño de la muestra de cada estrato.

Nivel socioeconomico	Nº de usuarios	W_h	n_h
A	2500		
B	1500		
C	1000		
Total	5000	1	50

Ejemplo

Estrato	Nh	Sh	Ch
1	162	28.96	20
2	132	20.36	30
3	61	8.54	40

Se disponemos de un presupuesto total de 3500 y costo fijo de 2500.

¿ Calcular el tamaño de muestra? Y el tamaño de cada estrato

B. Si V es fijo:

$$V_o = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} (1 - f_h)$$

$$n = \frac{\left(\sum_{h=1}^{L} W_h S_h \sqrt{C_h} \right) \left[\sum_{h=1}^{L} \frac{W_h S_h}{\sqrt{C_h}} \right]}{v(\bar{y}_{st}) + \frac{1}{N} \sum_{h=1}^{L} W_h S_h^2}$$

MUESTREO ESTRATIFICADO PARA PROPORCIONES

$$Y_{hi} = \begin{cases} 1 & \text{Si la unidad } i\text{-ésima del estrato } h \text{ tiene la característica} \\ 0 & \text{De otro modo} \end{cases}$$

ESTIMACION DE PARAMETROS - PROPORCIONES

- **MEDIA**

$$p_{st} = \sum_{h=1}^{L} W_h p_h$$

- **VARIANZA**

$$v(p_{st}) = \sum_{h=1}^{L} W_h^2 v(p_h)$$

Siendo:

$$v(p_h) = (1 - f_h)\frac{p_h q_h}{n_h - 1}$$

EJEMPLO

En un estudio de consumo de energía se desea estimar la proporción de edificios públicos que, según análisis, operan supuestamente de manera eficiente en lo que se refiere al uso de energía. Se dividió el territorio en tres zonas, dos grandes áreas urbanas y una rural. Los resultados de la encuesta, para la eficiencia energética, se tienen en la tabla siguiente:

	ZONA 1	**ZONA 2**	**ZONA 3**
No. Edificios públicos en la ZONA	**250**	**400**	**350**
Tamaño de Muestra	**50**	**80**	**70**
No. Edificios con uso eficiente de energía en la muestra	**14**	**34**	**29**

a-)¿Qué tipo de asignación se utilizó en este estudio?

b-)Estime la proporción de todos los edificios públicos que operan en forma eficiente en lo que se refiere al consumo de energía.

COMPARACION DE EFICIENCIAS SEGÚN LOS DISTINTOS DE ASIGNACION

A partir de:

$$S^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi} - \overline{Y}\right)^2$$

$$S^2 = \sum_{h=1}^{L} W_h S_h^2 + \sum_{h=1}^{L} W_h (\overline{Y}_h - \overline{Y})^2$$

$$V_{MAS}(\overline{y}) \geq V(\overline{y}_{st})_p$$

Comparación de las precisiones de la asignación proporcional y la óptima

$$V(\overline{y}_{st})_p - V(\overline{y}_{st})_o = \frac{1}{n}\left(\sum_{h=1}^{L} W_h S_h^2 + \left(\sum_{h=1}^{L} W_h S_h\right)^2\right) = \frac{1}{n}\sum_{h=1}^{L} W_h (S_h - \overline{S})^2$$

$$V(\overline{y}_{st})_p \geq V(\overline{y}_{st})_o$$

Por lo tanto:

$$V(\overline{y})_{MAS} = V(\overline{y}_{st})_p + \frac{1}{n}\sum_{h=1}^{L} W_h (\overline{Y}_h - \overline{Y})^2 = V(\overline{y}_{st})_o + \frac{1}{n}\sum_{h=1}^{L} W_h (S_h - \overline{S})^2 + \frac{(1-f)}{n}\sum_{h=1}^{L} W_h (\overline{Y}_h - \overline{Y})^2$$