MUESTREO ESTRATIFICADO

TECNICAS DE MUESTREO II

Email:cgonzales@lamolina.edu.pe

CONSTRUCCION DE LOS ESTRATOS

CONSTRUCCION DE LOS ESTRATOS

¿ Cuál es la mejor característica para la construcción de los estratos?

¿ Cómo se determinan los límites entre los estratos?

¿Cuántos estratos debería haber?

- 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:

- Conocer alguna característica de los hogares en Lima
- Estimar el consumo de energía eléctrica.

Disponibilidad de marcos.

Ejemplo:

En una encuesta de hogares:

Se utilizan planos catastrales para las zonas urbanas antiguas (un estrato), se usan fotografías aéreas para zonas rurales (otro estrato) y las áreas de posible nueva urbanización (otro estrato) se delimitan como otro marco; se muestrean áreas y se investigan las nuevas urbanizaciones (muestreo en etapas o conglomerados).

Costo de localizar y levantar la información de las unidades.

Ejemplo: en una encuesta de unidades agropecuarias.

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

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

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**

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

- **VARIANZA**

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

Siendo:

$$v(\overline{y}_h) = \left(1 - f_h\right) \frac{\hat{S}_h^2}{n_h} \qquad \hat{S}_h^2 = \sum_{i=1}^{n_h} \frac{\left(y_{hi} - \overline{y}_h\right)^2}{n_h - 1}$$

- **ERROR ESTANDAR**

$$\hat{S}_{\overline{y}_{st}} = \sqrt{v\left(\overline{y}_{st}\right)}$$

LIMITES DE CONFIANZA

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

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

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

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

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 $\overline{Y}$

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

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

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

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

AFIJACIÓN DEL TAMAÑO DE MUESTRA EN UNA POBLACIÓN ESTRATIFICADA ASUMIENDO UNA FUNCIÓN COSTO

Si el costo para obtener información de una unidad en el estrato *h*-ésimo es C_h, el costo total será:

$$C = C_0 + \sum_{h=1}^{L} C_h n_h$$

C_0 : es *costo administrativo*.

C_h:*Costo correspondiente al estrato h*

C: Costo total

La minimización de la varianza del estimador con costo fijo o viceversa, produce la *asignación óptima* que es:

$$n_h = n \frac{W_h S_h \Big/ \sqrt{C_h}}{\sum_{h=1}^{L} W_h S_h \Big/ \sqrt{C_h}}$$

$$n_h \propto \frac{N_h S_h}{\sqrt{C_h}}$$

TAMAÑO DE MUESTRA

A. Si el costo es fijo:

$$C = C_0 + \sum_{h=1}^{L} C_h n_h$$

$$n = \frac{\left(C - C_0\right) \sum_{h=1}^{L} \frac{N_h S_h}{\sqrt{C_h}}}{\sum_{h=1}^{L} N_h S_h \sqrt{C_h}}$$

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:

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

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

Ejemplo

Asumiendo Vo =12,25

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

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

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:

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

Asignación Optima

A: costo fijo:

$$n_h = n \frac{N_h \sqrt{\frac{P_h Q_h}{c_h}}}{\sum_{h=1}^{L} N_h \sqrt{\frac{P_h Q_h}{c_h}}}$$

$$n = \frac{\left(C - C_0\right) \sum_{h=1}^{L} N_h \sqrt{\frac{P_h Q_h}{C_h}}}{\sum_{h=1}^{L} N_h \sqrt{P_h Q_h} \sqrt{C_h}}$$

B. Si V es fijo:

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

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

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.

Ejemplo

Un especialista propone tomar una muestra aleatoria estratificada de una población que ha sido dividida en dos estratos; espera que sus costos de trabajo de campo tendrán la forma: $\sum c_h n_h$ sus estimaciones preliminares sobre los valores principalmente para los dos estratos son:

Estrato	W_h	s_h	C_h
I	0.4	10	4
II	0.6	20	9

a. Determinar los valores n_1/n, n_2/n que minimizan el costo de trabajo de campo de la investigación asumiendo una varianza una varianza predeterminada
b. Encontrar el tamaño de muestra requerido para que una asignación óptima se pueda lograr una $V(\bar{y}_{st}) = 1$. Ignore el factor de corrección.
c. Cuál será el costo total de trabajo de campo que se espera incurrir para la investigación.

Y: variable objetivo

X: variable complementaria que se utiliza para estratificar

$$V(\overline{Y}_{st}) = \frac{S^2}{n}\left[\frac{\rho^2}{L^2} + \left(1 - \rho^2\right)\right]$$

Ejemplo

- Sea: ρ=0,90
- C=1000L +10n
- Presupuesto= 10000

¿Cuál el número de estratos?

L	n	V(Yst)
1	900	0.0011111111 S^2
2	800	0.0004906250S^2
3	700	0.0004000000S^2
4	600	0.0004010417S^2
5	500	0.0004448000S^2
6	400	0.0005312500S^2
7	300	0.0006884354S^2
8	200	0.0010132813S^2
9	100	0.0020000000S^2

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

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

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

Por lo tanto:

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

TECNICA DE POST-ESTRATIFICACION

Estrato	N_h	W_h	M.A.S Clasificados después.				n_h	$\bar{y}_h$	$\hat{S}^2_h$
1	N_1	W_1	Y_{11}	Y_{12}	..	Y_{1n1}	n_1	$\bar{y}_1$	$\hat{S}^2_1$
2	N_2	W_2	Y_{21}	Y_{22}		Y_{2n2}	n_2	$\bar{y}_2$	$\hat{S}^2_2$
.									
.									
L	N_L	W_L	Y_{L1}	Y_{L2}		Y_{LnL}	n_L	$\bar{y}_L$	$\hat{S}^2_L$

PROMEDIO

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

VARIANZA

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

EFECTOS DE LAS DESVIACIONES A PARTIR DE LA ASIGNACION OPTIMA

$$n_h' = \frac{n\left(W_h S_h\right)}{\sum W_h S_h}$$

$$V_{\min}\left(\overline{y}_{st}\right) = \frac{1}{n}\left(\sum W_h S_h\right)^2 - \frac{1}{N}\sum W_h S_h^2$$

Por lo tanto:

$$\frac{V\left(\overline{y}_{st}\right) - V_{\min}\left(\overline{y}_{st}\right)}{V_{\min}\left(\overline{y}_{st}\right)} = \frac{1}{n}\sum_{h=1}^{L}\frac{\left(n_h' - \hat{n}_h\right)^2}{\hat{n}_h}$$

Efecto de las desviaciones de la asignación óptima

Estrato	$n_h^{'}$	$\hat{n}_h$	$\frac{\lvert n_h^{'} - \hat{n}_h \rvert}{\hat{n}_h}$	$\frac{\left(n_h^{'} - \hat{n}_h\right)^2}{\hat{n}_h}$
1	400	300	0.33	33.33
2	200	240	0.17	6.67
3	80	140	0.43	25.71
	680	680		65.71

Incremento real:

Para poblaciones desproporcionadas

Estrato	N_h	n'_h optimo	$\hat{n}_h$ valoral	$\hat{n}_h$ proporcional
1	52710	444	321	488
2	1190	43	129	11
3	50	13	50	1
Total	53950	500	500	500

EJERCICIO

Suponga que durante el mes de febrero del año 2007 un determinado establecimiento ha tenido 100 clientes y que dispone de la información sobre las compras en nuevos soles que ha realizado cada uno de ellos; Explica con un caso concreto (genere los datos) el procedimiento que seguirías para obtener una muestra de tamaño 10

a. Mediante un muestreo aleatorio simple

b. Mediante un muestreo sistemático con arranque aleatorio.

c. Mediante un muestreo con probabilidades desiguales en el que se da tanta mayor probabilidad de salida a los individuos cuánto mayor sea su volumen de compras (probabilidades proporcionales al tamaño).

d. Que ventajas principales aporta el muestreo sistemático con arranque aleatorio sobre el muestreo aleatorio simple?

EJERCICIO

En el ejemplo creado anteriormente, genera una muestra de 20 clientes mediante muestreo estratificado; explica que criterio has seguido para determinar.

a. El número de estratos a considerar

b. Los límites de dichos estratos.

c. La asignación de la muestra por estratos