3. Muestreo con información auxiliar

3. Muestreo con información auxiliar

3.1 Introducción

$Y \longrightarrow$ Variable bajo estudio

$X \longrightarrow$ Variable que proporciona la información auxiliar

Muestra constituida por *n* pares: $(x_1, y_1), \ldots, (x_n, y_n)$

Bajo una fuerte relación lineal positiva $\left(r_{xy} > \frac{1}{2} \right)$

y dependiendo de la relación entre ambas variables utilizaremos:

- Estimadores de razón $(y = bx)$
- Estimadores de regresión $(y = a + bx)$
- Estimadores de diferencia $(y = a + x)$

3. Muestreo con información auxiliar

3.2 Estimación de razón

Se define la razón como el cociente:

$$R = \frac{\tau_y}{\tau_x} \longrightarrow R = \frac{\mu_Y}{\mu_X}$$

$$\tau_y = N\mu_y$$

$$\tau_x = N\mu_x$$

Entonces si se conocen los valores de la media y del total de *X* sólo hay que estimar el valor de *R (r):*

$$\hat{\tau}_y = r\tau_x$$

$$\hat{\mu}_y = r\mu_x$$

• ESTIMADOR DE LA RAZÓN: $r = \frac{\overline{y}}{\overline{x}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$

• VARIANZA ESTIMADA DE *r:* $\hat{V}(r) = \frac{1}{\mu_x^2}\frac{S_r^2}{n}\left(\frac{N-n}{N}\right)$ $\quad S_r^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - rx_i)^2$

3

3. Muestreo con información auxiliar

3.2 Estimación de razón

3.2.1 Estimación de la media y el total poblacionales

Entre *X* e *Y* existe una alta correlación lineal positiva y que el modelo lineal pasa por el origen.

• ESTIMADOR DE LA MEDIA: $\hat{\mu}_y = r\mu_x$

•VARIANZA ESTIMADA DEL ESTIMADOR: $\hat{V}(\hat{\mu}_y) = \mu_x^2\hat{V}(r) = \frac{S_r^2}{n}\left(\frac{N-n}{N}\right)$

• ESTIMADOR DEL TOTAL: $\hat{\tau}_y = r\tau_x$

•VARIANZA ESTIMADA DEL ESTIMADOR:

$$\hat{V}(\hat{\tau}_y) = \tau_x^2\hat{V}(r) = \frac{\tau_x^2}{\mu_x^2}\frac{S_r^2}{n}\left(\frac{N-n}{N}\right) = N^2\frac{S_r^2}{n}\left(\frac{N-n}{N}\right)$$

4

3. Muestreo con información auxiliar

3.2 Estimación de razón

3.2.1 Estimación de la media y el total poblacionales

	Población finita	Total poblacional, τ_Y	Media poblacional, μ_Y	Varianza poblacional, σ_Y^2
①$_4$ ②$_8$ ③$_{14}$ ⓫$_{56}$ ⓬$_{62}$ ⓭$_{66}$	N=6	$\tau_Y = 1+2+3+11+12+13 = 42$	$\mu_Y = \frac{42}{6} = 7$	25,6667
		Total poblacional, τ_X	Media poblacional, μ_X	Varianza poblacional, σ_X^2
		$\tau_X = 4+8+14+56+62+66 = 210$	$\mu_X = \frac{210}{6} = 35$	710,3333

$$R = \frac{\tau_Y}{\tau_X} = \frac{\mu_Y}{\mu_X} = \frac{42}{210} = \frac{7}{35} = \frac{1}{5} = 0,2$$

5

5

3. Muestreo con información auxiliar

3.2 Estimación de razón

3.2.1 Estimación de la media y el total poblacionales

	MUESTRAS (n=4)	Media muestral, $\bar{x}$	Media muestral, $\bar{y}$	$r = \frac{\bar{y}}{\bar{x}}$	$\hat{\mu}_Y = r\mu_X$
1	③$_{14}$ ⓫$_{56}$ ⓬$_{62}$ ⓭$_{66}$	$\frac{14+56+62+66}{4} = 49,5$	$\frac{3+11+12+13}{4} = 9,75$	$\frac{9,75}{49,5} = 0,197$	6,8939
2	②$_8$ ⓫$_{56}$ ⓬$_{62}$ ⓭$_{66}$	48	9,5	0,1979	6,9271
3	②$_8$ ③$_{14}$ ⓬$_{62}$ ⓭$_{66}$	37,5	7,5	0,2000	7,0000
4	②$_8$ ③$_{14}$ ⓫$_{56}$ ⓭$_{66}$	36	7,25	0,2014	7,0486
5	②$_8$ ③$_{14}$ ⓫$_{56}$ ⓬$_{62}$	35	7	0,2000	7,0000
6	①$_4$ ⓫$_{56}$ ⓬$_{62}$ ⓭$_{66}$	47	9,25	0,1968	6,8883
7	①$_4$ ③$_{14}$ ⓬$_{62}$ ⓭$_{66}$	36,5	7,25	0,1986	6,9521
8	①$_4$ ③$_{14}$ ⓫$_{56}$ ⓭$_{66}$	35	7	0,2000	7,0000
9	①$_4$ ③$_{14}$ ⓫$_{56}$ ⓬$_{62}$	34	6,75	0,1985	6,9485
10	①$_4$ ②$_8$ ⓬$_{62}$ ⓭$_{66}$	35	7	0,2000	7,0000
11	①$_4$ ②$_8$ ⓫$_{56}$ ⓭$_{66}$	33,5	6,75	0,2015	7,0522
12	①$_4$ ②$_8$ ⓫$_{56}$ ⓬$_{62}$	32,5	6,5	0,2000	7,0000
13	①$_4$ ②$_8$ ③$_{14}$ ⓭$_{66}$	23	4,75	0,2065	7,2283
14	①$_4$ ②$_8$ ③$_{14}$ ⓬$_{62}$	22	4,5	0,2045	7,1591
15	①$_4$ ②$_8$ ③$_{14}$ ⓫$_{56}$	20,5	4,25	0,2073	7,2561

6

6

3. Muestreo con información auxiliar

3.2 Estimación de razón

3.2.1 Estimación de la media y el total poblacionales

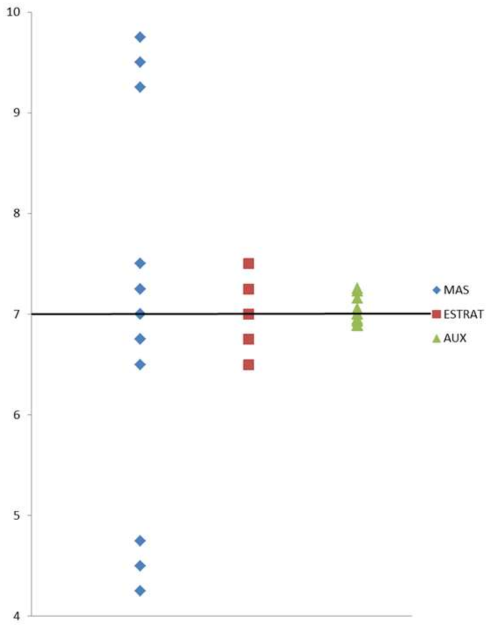
MUESTRAS (n=4)	Media muestral, $\bar{x}$	Media muestral, $\bar{y}$	$r = \frac{\bar{y}}{\bar{x}}$	$\hat{\mu}_Y = r\mu_X$
MEDIA:	$E\left[\bar{x}\right] = \frac{525}{15} = 35$	$E\left[\bar{y}\right] = \frac{105}{15} = 7$	$E\left[r\right] = 0,20067$	$E\left[\hat{\mu}_Y\right] = 7,0236$
VARIANZA:	$V\left[\bar{x}\right] = 71,0333$	$V\left[\bar{y}\right] = 2,5667$	$V\left[r\right] = 0,00001$	$V\left[\hat{\mu}_Y\right] = 0,0116$

MUESTRAS n=4 (n_1=2, n_2=2)		$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_{st}$
②③	⓬⓭	$\frac{2+3}{2} = 2,5$	$\frac{12+13}{2} = 12,5$	$\frac{(3\times 2,5)+(3\times 12,5)}{6} = 7,5$
②③	⓫⓭	2,5	12	7,25
②③	⓫⓬	2,5	11,5	7
①③	⓬⓭	2	12,5	7,25
①③	⓫⓭	2	12	7
①③	⓫⓬	2	11,5	6,75
①②	⓬⓭	1,5	12,5	7
①②	⓫⓭	1,5	12	6,75
①②	⓫⓬	1,5	11,5	6,5
TOTAL:		18	108	63
MEDIA:		$E\left[\bar{y}_1\right] = \frac{18}{9} = 2$	$E\left[\bar{y}_2\right] = \frac{108}{9} = 12$	$E\left[\bar{y}_{st}\right] = \frac{63}{9} = 7$
VARIANZA:		$V\left[\bar{y}_1\right] = 0,1667$	$V\left[\bar{y}_2\right] = 0,1667$	$V\left[\bar{y}_{st}\right] = 0,0833$

7

7

3. Muestreo con información auxiliar



8

8

3. Muestreo con información auxiliar

3.2 Estimación de razón

3.2.2 Determinación del tamaño muestral

Tamaño muestral mínimo para que la estimación de la razón, de la media y del total no supere una cota de error de magnitud B

$$n = \frac{N\sigma_r^2}{\sigma_r^2 + ND} \qquad D = \begin{cases} \dfrac{B^2\mu_x^2}{4} & \text{para estimar la razón} \\ \dfrac{B^2}{4} & \text{para estimar la media} \\ \dfrac{B^2}{4N^2} & \text{para estimar el total} \end{cases}$$

- *N* debe ser conocido o estimado. Si N es infinito $n = \dfrac{\sigma_r^2}{D}$
- σ_r^2 se estima utilizando una muestra previa (tamaño *n'*): $\hat{\sigma}_r^2 = S_r^2$
- $\hat{\mu}_x^2 = \bar{x}^2$

9

3. Muestreo con información auxiliar

3.3 Estimación de regresión

Entre *X* e *Y* existe una alta correlación lineal positiva y el modelo lineal no pasa por el origen.

Modelo lineal simple

$$y = a + bx$$

Método de mínimos cuadrados

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$y = \left(\bar{y} - \hat{b}\bar{x}\right) + \hat{b}x = \bar{y} + \hat{b}\left(x - \bar{x}\right)$$

Recuerde: $s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \frac{1}{n}\sum_{i=1}^{n}x_i y_i - \bar{x}\,\bar{y}$

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

10

3. Muestreo con información auxiliar

3.3 Estimación de regresión $y = \left(\bar{y} - \hat{b}\bar{x}\right) + \hat{b}x = \bar{y} + \hat{b}\left(x - \bar{x}\right)$

3.3.1 Estimación de la media y el total poblacionales

- ESTIMADOR DE LA MEDIA: $\hat{\mu}_{yL} = \hat{a} + \hat{b}\mu_x = \bar{y} + \hat{b}(\mu_x - \bar{x})$
- VARIANZA ESTIMADA DEL ESTIMADOR: $\hat{V}(\hat{\mu}_{yL}) = \frac{S_L^2}{n}\left(\frac{N-n}{N}\right)$

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

$$= \frac{n}{n-2}\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right) = \frac{n}{n-2}s_y^2\left(1 - r_{xy}^2\right)$$

- ESTIMADOR DEL TOTAL: $\hat{\tau}_{yL} = N\hat{\mu}_{yL}$
- VARIANZA ESTIMADA DEL ESTIMADOR: $\hat{V}(\hat{\tau}_{yL}) = N^2\hat{V}(\hat{\mu}_{yL})$

11

11

3. Muestreo con información auxiliar

3.3 Estimación de regresión

3.3.2 Determinación del tamaño muestral

Tamaño muestral mínimo para que la estimación de la media y del total no supere una cota de error de magnitud B

$$n = \frac{N\sigma_L^2}{\sigma_L^2 + ND} \qquad D = \begin{cases} \frac{B^2}{4} & \text{para estimar la media} \\ \frac{B^2}{4N^2} & \text{para estimar el total} \end{cases}$$

- σ_L^2 se estima utilizando una muestra previa (tamaño *n'*): $\hat{\sigma}_L^2 = S_L^2$
- *N* debe ser conocido o estimado. Si N es infinito $n = \frac{\sigma_L^2}{D}$

12

12

3. Muestreo con información auxiliar

3.4 Estimación de diferencia

Entre *X* e *Y* existe una alta correlación lineal positiva y la pendiente del modelo es uno.

$$y = \left(\overline{y} - \hat{b}\overline{x}\right) + \hat{b}x = \overline{y} + \hat{b}\left(x - \overline{x}\right)$$

3.4.1 Estimación de la media y el total poblacionales

- ESTIMADOR DE LA MEDIA: $\hat{\mu}_{yD} = \overline{y} + \left(\mu_x - \overline{x}\right) = \mu_x + \overline{d}$

  donde $\overline{d} = \overline{y} - \overline{x}$

- VARIANZA ESTIMADA DEL ESTIMADOR: $\hat{V}\left(\hat{\mu}_{yD}\right) = \frac{S_D^2}{n}\left(\frac{N-n}{N}\right)$

  donde $S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \left(x_i + \overline{d}\right)\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2$ y $d_i = y_i - x_i$

  $S_D^2 = cuasivarianza\ de\ los\ d_i$

- ESTIMADOR DEL TOTAL: $\hat{\tau}_{yD} = N\hat{\mu}_{yD}$

- VARIANZA ESTIMADA DEL ESTIMADOR: $\hat{V}\left(\hat{\tau}_{yD}\right) = N^2\hat{V}\left(\hat{\mu}_{yD}\right)$

13

13

3. Muestreo con información auxiliar

3.4 Estimación de diferencia

3.4.2 Determinación del tamaño muestral

Tamaño muestral mínimo para que la estimación de la media y del total no supere una cota de error de magnitud B

$$n = \frac{N\sigma_D^2}{\sigma_D^2 + ND} \qquad D = \begin{cases} \frac{B^2}{4} & \text{para estimar la media} \\ \frac{B^2}{4N^2} & \text{para estimar el total} \end{cases}$$

- σ_D^2 se estima utilizando una muestra previa (tamaño *n'*): $\hat{\sigma}_D^2 = S_D^2$
- *N* debe ser conocido o estimado. Si N es infinito $n = \frac{\sigma_D^2}{D}$

14

14