

# Técnicas de Muestreo I

Patricia Isabel Romero Mares

Departamento de Probabilidad y Estadística  
IIMAS UNAM

septiembre 2015

## Estimadores de Razón (bajo m.a.s.)

# Estimadores de Razón

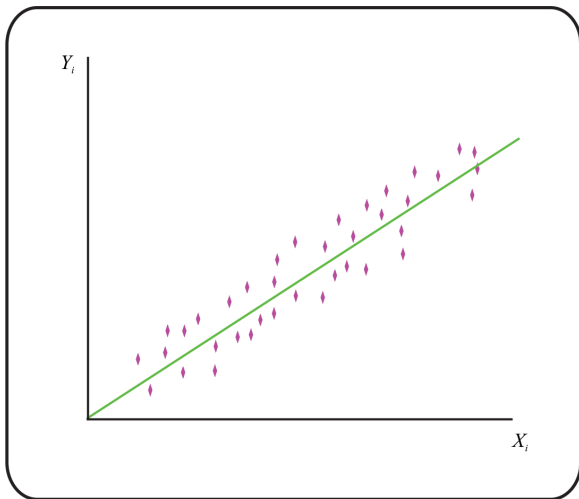
Se hace uso de información auxiliar.

Suponga que para cada unidad muestral, además de obtener información acerca de una variable  $Y$ , se obtiene información de una variable  $X$ , y se sabe que  $Y$  y  $X$  están correlacionadas.

El estimador de razón dará una estimación de  $Y$  con más precisión que el estimador usual  $\hat{Y}_{m.a.s.}$ .

## Estimador de razón

**Caso 1.** Se quiere estimar  $Y$  ó  $\bar{Y}$  bajo el supuesto que  $Y_i \propto X_i$  y se conoce  $X$  ó  $\bar{X}$ .



$$R = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N X_i} = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}}$$

entonces,

$$Y = RX \quad \text{ó} \quad \bar{Y} = R\bar{X}$$

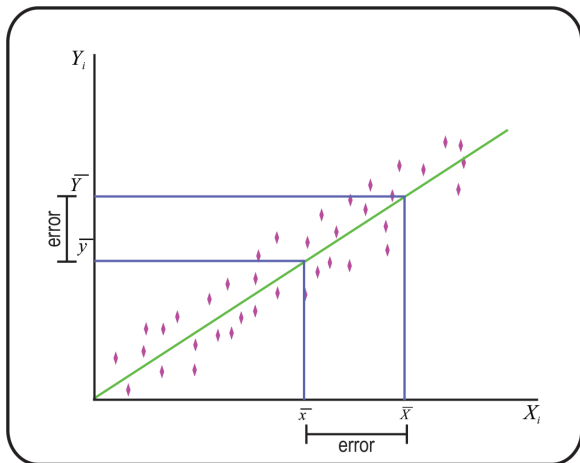
Un estimador basado en una m.a.s. de  $n$  elementos es:

$$\hat{Y} = \hat{R}X \quad \hat{\bar{Y}} = \hat{R}\bar{X}$$

donde,

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$$

## Estimador de razón



$$Y_i \propto X_i$$

$$Y_i \cong RX_i$$

$$Y_i = RX_i + \varepsilon_i$$

$$\varepsilon_i = Y_i - RX_i$$

Como conocemos  $\bar{X}$ , conocemos la tasa de error  $\frac{\bar{X}}{\bar{x}}$ , entonces

$$\hat{Y} = \hat{R}\bar{X} = \frac{\bar{y}}{\bar{x}}\bar{X} = \bar{y} \underbrace{\left(\frac{\bar{X}}{\bar{x}}\right)}_{\text{ajuste}}$$

Si, por ejemplo,

$$\frac{\bar{X}}{\bar{x}} = 1.2 \Rightarrow \bar{X} = 1.2\bar{x} \text{ estamos subestimando, i.e. } \bar{x} < \bar{X}$$

## Ejemplo

Suponga que se tiene una m.a.s. de 49 ciudades de un total de 196 de cierta región del país, de las que se conoce el número de habitantes en el año 2010 y se quiere estimar el total de habitantes en la región en 2014.

Se conoce el total de habitantes en 2010,  $X = 22919$

$$\sum_{i=1}^{49} x_i = 5054; \sum_{i=1}^{49} y_i = 6262$$

$$\hat{Y}_R = \hat{R}X = \frac{\sum_{i=1}^{49} y_i}{\sum_{i=1}^{49} x_i} X = \frac{6262}{5054} (22919) = 28397$$

Si consideramos el estimador usual del total bajo m.a.s.:

$$\hat{Y}_{m.a.s.} = N\bar{y} = 196 \left( \frac{6262}{49} \right) = 25048$$

El total real en el año 2014 es 29351.



**Caso 2.** Se quiere estimar  $R$  (estimador de la razón)

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum y_i}{\sum x_i}$$

Por ejemplo, en una encuesta de familias, se mide el ingreso total familiar ( $y_i$ ) y el número de miembros de la familia ( $x_i$ ), entonces, se podría obtener el ingreso *per cápita*:

$$\hat{R} = \frac{\sum y_i}{\sum x_i}.$$

Ejemplos de este tipo surgen cuando la unidad de muestreo (en el ejemplo, la familia), comprende un conjunto de elementos (miembros de la familia) y nuestro interés es estimar la **media por elemento**.

También cuando se quiere estimar la proporción de cierta característica en relación al total de todas las características.

Por ejemplo:

$$\% \text{ votos al partido } q = \frac{\text{total de votos al partido } q}{\text{total de votos}}$$

donde

$$\text{total de votos} = \text{votos al partido 1} + \text{votos al partido 2} + \dots$$

Tomado de Raj y Chandhok (1998).

$\hat{R}$  es consistente para  $R$  en el sentido de que  $\hat{R} = R$  cuando el tamaño de muestra es  $N$ .

$\hat{R}$  es un estimador sesgado,  $E(\hat{R}) \neq R$

$$ECM(\hat{R}) = V(\hat{R}) + [B(\hat{R})]^2$$

$$B(\hat{R}) = E(\hat{R} - R) \approx R CV(\bar{x}) [CV(\bar{x}) - \rho CV(\bar{y})]$$

Además, se demuestra que

$$|B(\hat{R})| = |E(\hat{R}) - R| \leq CV(\bar{x}) \sqrt{V(\hat{R})}$$

O bien,

$$\frac{|E(\hat{R}) - R|}{\sqrt{V(\hat{R})}} \leq CV(\bar{x}) = \frac{\sqrt{V(\bar{x})}}{\bar{x}}$$

En estudios de simulación se ha visto que:

$\frac{ B(\hat{R}) }{\sqrt{V(\hat{R})}}$	% $[R \in I_{95\%}]$
0	0.95
0.01	0.95
0.10	0.9481
0.30	0.9396
0.50	0.9210
1	0.83

## Estimador de razón

Si  $n$  es **grande** y/o el  $CV(\bar{x}) \leq 0.1$  el sesgo es despreciable y se toma  $V(\hat{R})$  en lugar de  $ECM(\hat{R})$ . Además, la distribución muestral de  $\hat{R}$  se aproxima a la normal.

$$V(\hat{R}) = \frac{1}{\bar{X}^2} \left( \frac{1}{n} - \frac{1}{N} \right) [S_Y^2 - 2RS_{XY} + R^2S_X^2]$$

$$S_X^2 = \frac{\sum^N (X_i - \bar{X})^2}{N - 1}$$

$$S_Y^2 = \frac{\sum^N (Y_i - \bar{Y})^2}{N - 1}$$

$$S_{XY} = \frac{\sum^N (X_i - \bar{X})(Y_i - \bar{Y})}{N - 1}$$

y su estimador es:

$$\hat{V}(\hat{R}) = \frac{1}{\bar{x}^2} \left( \frac{1}{n} - \frac{1}{N} \right) [\hat{S}_Y^2 - 2\hat{R}\hat{S}_{XY} + \hat{R}^2\hat{S}_X^2]$$

Sin embargo, tiene una expresión más operativa (Cochran)

$$\begin{aligned} V(\hat{R}) &= \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{\bar{X}^2} \sum_{i=1}^N \frac{(Y_i - RX_i)^2}{N-1} \\ &= \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{\bar{X}^2} V(\varepsilon_i) \text{ con } \varepsilon_i = Y_i - RX_i \end{aligned}$$

La varianza depende de la varianza de los errores. Si hay buena proporcionalidad entre  $X$  y  $Y$ , es decir, si  $Y_i \cong RX_i$ , la varianza del estimador de  $R$  es pequeña.

Se estima con:

$$\hat{V}(\hat{R}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{\bar{x}^2} \sum_{i=1}^n \frac{(y_i - \hat{R}x_i)^2}{n-1}$$

## Resumen Estimador de razón

Para estimar la **razón** poblacional:

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum y_i}{\sum x_i}$$

$$V(\hat{R}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{\bar{X}^2} \sum_{i=1}^N \frac{(Y_i - RX_i)^2}{N-1}$$

$$\hat{V}(\hat{R}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{1}{\bar{x}^2} \sum_{i=1}^n \frac{(y_i - \hat{R}x_i)^2}{n-1}$$

Si  $n$  es suficientemente grande para que aplique la aproximación Normal, el intervalo aproximado de  $(1 - \alpha) \times 100\%$  de confianza para  $R$  es:

$$\hat{R} \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{R})}$$

Para estimar el **total** poblacional (con  $X$  fijo y conocido):

$$\hat{Y} = \hat{R}X$$

$$V(\hat{Y}) = X^2 V(\hat{R}) = N^2 \left[ \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{\sum_{i=1}^N (Y_i - RX_i)^2}{N-1} \right]$$

$$\hat{V}(\hat{Y}) = N^2 \left[ \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{R}x_i)^2}{n-1} \right]$$

Si  $n$  es suficientemente grande para que aplique la aproximación Normal, el intervalo aproximado de  $(1 - \alpha) \times 100\%$  de confianza para  $Y$  es:

$$\hat{Y} \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{Y})}$$



Para estimar la **media** poblacional (con  $\bar{X}$  fijo y conocido):

$$\hat{\bar{Y}} = \hat{R}\bar{X}$$

$$V(\hat{\bar{Y}}) = \bar{X}^2 V(\hat{R}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{\sum_{i=1}^N (Y_i - RX_i)^2}{N-1}$$

$$\hat{V}(\hat{\bar{Y}}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{R}x_i)^2}{n-1}$$

Si  $n$  es suficientemente grande para que aplique la aproximación Normal, el intervalo aproximado de  $(1 - \alpha) \times 100\%$  de confianza para  $\bar{Y}$  es:

$$\hat{\bar{Y}} \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\bar{Y}})}$$

## Comparación estimador de razón con el usual de m.a.s.

$$\hat{Y}_{m.a.s.} = N\bar{y} \quad \hat{Y}_R = \hat{R}X$$

$$\begin{aligned} V(\hat{Y}_{m.a.s.}) &= N^2 \left(1 - \frac{n}{N}\right) \frac{S_Y^2}{n} \\ V(\hat{Y}_R) &= \frac{N^2}{n} \left(1 - \frac{n}{N}\right) (S_Y^2 - 2RS_{XY} + R^2 S_X^2) \\ &= \frac{N^2}{n} \left(1 - \frac{n}{N}\right) (S_Y^2 - 2R\rho S_X S_Y + R^2 S_X^2) \end{aligned}$$

donde,

$$\rho = \frac{S_{XY}}{S_X S_Y}$$

## Comparación estimador de razón con el usual de m.a.s.

El estimador de razón es más preciso que el estimador usual de m.a.s. si

$$V(\hat{Y}_R) < V(\hat{Y}_{m.a.s.})$$

Si

$$S_Y^2 + R^2 S_X^2 - 2R\rho S_X S_Y < S_Y^2 \Rightarrow$$

$$R^2 S_X^2 - 2R\rho S_X S_Y < 0 \Rightarrow$$

$$\rho > \frac{R^2 S_X^2}{2R S_X S_Y} \Rightarrow$$

$$\rho > \frac{R S_X}{2 S_Y} = \frac{CV(X)}{2 CV(Y)}$$

## Tamaño de muestra

Si se especifica una  $\delta$  para el error de estimación en  $\bar{Y}$ , esto es,

$$P \left[ \left| \hat{Y}_R - \bar{Y} \right| < \delta \right] = 1 - \alpha$$

Se obtendrá que el tamaño de muestra adecuado, si  $N$  es grande, es:

$$n = \frac{z_{1-\alpha/2}^2 S_\varepsilon^2}{\delta^2}$$

Donde,

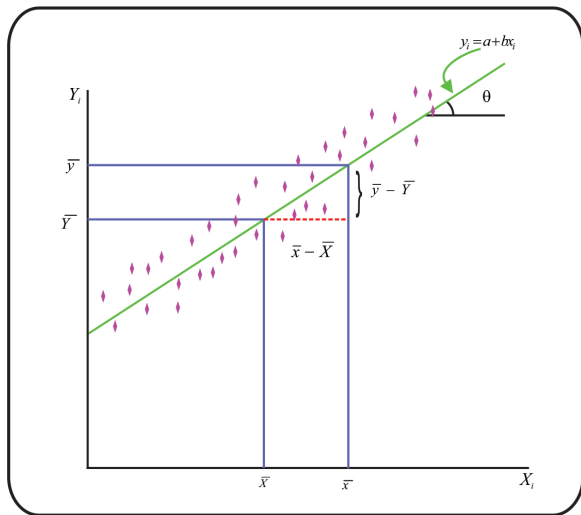
$$S_\varepsilon^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - RX_i)^2$$

Recordemos que en el caso del estimador usual del m.a.s.

$$n = \frac{z_{1-\alpha/2}^2 S_Y^2}{\delta^2}$$

## Estimadores de Regresión (bajo m.a.s.)

# Estimadores de regresión



Variable de interés  $Y_i$

Variable auxiliar  $X_i$

$$b = \tan \theta = \frac{\text{cateto opuesto}}{\text{cateto adyacente}}$$

$$b = \frac{\bar{y} - \bar{Y}}{\bar{x} - \bar{X}}$$

$$b(\bar{x} - \bar{X}) = \bar{y} - \bar{Y}$$

El estimador de regresión de la **media** poblacional es:

$$\hat{Y}_{reg} = \bar{y} - \hat{b}(\bar{x} - \bar{X}) = \bar{y} + \hat{b}(\bar{X} - \bar{x}) = \hat{Y}_{mas} + \hat{b}(\bar{X} - \hat{X}_{mas})$$

El estimador de regresión del **total** poblacional es:

$$N\hat{Y} = N\bar{y} + N\hat{b}(\bar{X} - \bar{x})$$

$$\hat{Y}_{reg} = \hat{Y}_{m.a.s.} + \hat{b}(X - \hat{X}_{m.a.s.})$$

## estimador de regresión

Donde  $\hat{b}$  es la pendiente

$$\hat{b} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\hat{S}_{XY}}{\hat{S}_X^2}$$

además, es el estimador de mínimos cuadrados. Minimiza la  $V(\hat{Y}_{reg})$ . Es difícil encontrar expresiones exactas para la varianza o los ECM de estos estimadores (son sesgados). Sin embargo, si  $n$  es grande:

$$V(\hat{Y}) = \frac{S_Y^2}{n} (1 - \rho^2) \left(1 - \frac{n}{N}\right)$$

donde  $\rho$  es el coeficiente de correlación entre  $X$  y  $Y$ .

$$\rho = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2\right]^{1/2}}$$



## Resumen estimador de regresión

Para estimar la **media** poblacional, con  $\bar{X}$  conocido:

$$\hat{\bar{Y}}_{reg} = \bar{y} + \hat{b}(\bar{X} - \bar{x})$$

$$V(\hat{\bar{Y}}_{reg}) = \frac{S_Y^2}{n} (1 - \rho^2) \left(1 - \frac{n}{N}\right)$$

$$\hat{V}(\hat{\bar{Y}}_{reg}) = \frac{\hat{S}_Y^2}{n} (1 - \hat{\rho}^2) \left(1 - \frac{n}{N}\right)$$

Si  $n$  es suficientemente grande para que aplique la aproximación Normal, el intervalo aproximado de  $(1 - \alpha) \times 100\%$  de confianza para  $\bar{Y}$  es:

$$\hat{\bar{Y}}_{reg} \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\bar{Y}}_{reg})}$$

## Resumen estimador de regresión

Para estimar el **total** poblacional, con  $X$  conocido:

$$\hat{Y}_{reg} = N\bar{y} + \hat{b}(X - N\bar{x})$$

$$V(\hat{Y}_{reg}) = N^2 \frac{S_Y^2}{n} (1 - \rho^2) \left(1 - \frac{n}{N}\right)$$

$$\hat{V}(\hat{Y}_{reg}) = N^2 \frac{\hat{S}_Y^2}{n} (1 - \hat{\rho}^2) \left(1 - \frac{n}{N}\right)$$

Si  $n$  es suficientemente grande para que aplique la aproximación Normal, el intervalo aproximado de  $(1 - \alpha) \times 100\%$  de confianza para  $Y$  es:

$$\hat{Y}_{reg} \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{Y}_{reg})}$$

## Ejemplo estimador de regresión

Se examinó a 486 candidatos a ingresar a una escuela. De éstos se tomó una m.a.s. de 10 estudiantes a los que se les midió su calificación en Cálculo al final del primer semestre.

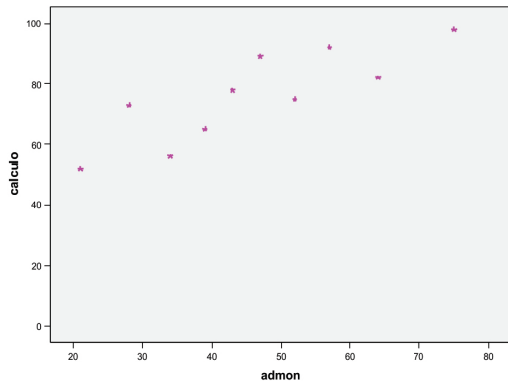
Se sabe que  $\bar{X} = 52$  para las 486 estudiantes.

Se desea estimar  $\bar{Y}$ , el promedio de calificación en Cálculo de todos los estudiantes al final del primer semestre.

## Ejemplo estimador de regresión

estudiante	Calificación admisión (X)	Calificación Cálculo (Y)
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

## Ejemplo estimador de regresión



## Ejemplo estimador de regresión

$$\begin{array}{lll} N = 486 & \bar{X} = 52 & \bar{x} = 46 \\ n = 10 & \bar{y} = 76 & \hat{S}_Y^2 = 228.444 \\ \hat{b} = 0.766 & \hat{\rho} = 0.84 & \end{array}$$

$$\begin{aligned} \hat{Y}_{reg} &= \bar{y} + \hat{b} (\bar{X} - \bar{x}) \\ &= 76 + (0.766) (52 - 46) \\ \hat{Y}_{reg} &= 80.596 \end{aligned}$$

$$\begin{aligned} \hat{V}(\hat{Y}_{reg}) &= \left(1 - \frac{n}{N}\right) \frac{\hat{S}_Y^2}{n} (1 - \hat{\rho}^2) \\ &= \left(1 - \frac{10}{486}\right) \frac{228.444}{10} (1 - 0.84^2) \\ &= (0.9794) (22.844) (0.2944) \\ &= 6.586 \end{aligned}$$

## Ejemplo estimador de regresión

Si hacemos la aproximación a normalidad, el Intervalo del 95% de confianza para  $\bar{Y}$  es:

$$\begin{aligned}\hat{Y}_{reg} &\pm 1.96\sqrt{6.586} \\ 80.596 &\pm 1.96(2.566) \\ 80.596 &\pm 5.029 \\ (75.567 &, 85.625)\end{aligned}$$

# Tamaño de muestra

Si se especifica una  $\delta$  para el error de estimación en  $\bar{Y}$ , esto es,

$$P \left[ \left| \hat{Y}_{reg} - \bar{Y} \right| < \delta \right] = 1 - \alpha$$

Se obtendrá que el tamaño de muestra adecuado, si  $N$  es grande, es:

$$n = \frac{z_{1-\alpha/2}^2 S_Y^2 (1 - \rho^2)}{\delta^2}$$

Si  $\rho$  es grande,  $n$  es pequeña.

Recordemos que en el caso del estimador usual del m.a.s.

$$n = \frac{z_{1-\alpha/2}^2 S_Y^2}{\delta^2}$$