

[10] De los 750 trabajadores de una fábrica, se conoce que el número medio de días anuales de ausencia del trabajo sin justificar para las mujeres (variable X) es 10 y para los hombres (variable Y) es 8. Se sabe que el error cometido al cuantificar la media de la variable X es 2500 y que la razón de la covarianza de X e Y a la varianza de X es 0,6.

- (a) (5 puntos) Determinar a partir de qué tamaño muestral el sesgo del estimador de la razón de X/Y es **despreciable** utilizando un muestreo sin reposición.
- (b) (5 puntos) ¿Qué método de estimación indirecta sería el más adecuado a utilizar sobre muestras de esta población? Justique.

Solución: Por enunciado del problema se tiene:

$$\bar{X} = 10, \quad \bar{Y} = 8, \quad \sigma_x^2 = 2500, \quad \frac{\sigma_{xy}}{\sigma_x^2} = 0,6$$

De la condición de que el sesgo relativo $|\frac{B(\hat{R})}{\sigma(\hat{R})}|$ sea mejor que un décimo se tiene que:

$$n \geq \frac{N \cdot 100 \cdot S_x^2}{N\bar{X}^2 + 100S_x^2} = 577$$

En caso de muestreo con reposición la misma condición de sesgo relativo menor que un décimo nos lleva a $n \geq 2500$, que sobrepasa el tamaño poblacional (por lo que no podría haber sesgo despreciable).

La recta de regresión de Y sobre X tiene ecuación $y - \bar{y} = \frac{\hat{S}_{xy}}{\hat{S}_x^2}(x - \bar{x}) \Rightarrow y - 8 = 0,6(x - 10) \Rightarrow y = 0,6x + 2$, lo que indica que la estimación por razón podría ser adecuada al no ser demasiado grande la ordenada en el origen. La estimación por regresión siempre es el método más adecuado. La pendiente de la recta no es unitaria, con lo que no es muy apropiada la estimación por diferencia.