

[30] Una muestra de 100 estudiantes es escogida usando un muestreo aleatorio simple sin reemplazo desde una población de 1000 estudiantes. Se está interesado en los resultados obtenidos por estos estudiantes en una prueba. Esta prueba sólo tiene dos posibles resultados, aprobado o reprobado. Los resultados obtenidos se presentan en la siguiente tabla.

	Hombres	Mujeres	Total
Aprobados	$n_{11} = 35$	$n_{12} = 25$	$n_{1.} = 60$
Reprobados	$n_{21} = 20$	$n_{22} = 20$	$n_{2.} = 40$
Total	$n_{.1} = 55$	$n_{.2} = 45$	$n = 100$

- (a) (7 puntos) Estime la tasa de aprobación de los hombres y de las mujeres.  
 (b) (7 puntos) Calcule el sesgo aproximado para estas estimaciones.  
 (c) (7 puntos) Estime el error cuadrático medio de estas tasas de aprobación.  
 (d) (8 puntos) Calcule los intervalos de confianza al 95 % para las tasas de aprobación de hombres y mujeres. ¿Qué puede decir de sus respectivas posiciones?

**Solución:** Es claro que las tasas de aprobación estarán dadas por:

$$r_H = \frac{35}{55} \approx 63,6 \% \quad r_M = \frac{25}{45} \approx 55,6 \%$$

Estas tasas corresponden a estimadores de razón. Debido a que la muestra es de tamaño 100, consideramos el tamaño de muestra  $n$  lo suficientemente grande. Así, el sesgo de la razón:  $r = \hat{Y}/\hat{X}$  estará dado por:

$$B(r) = R \left( \frac{S_x^2}{\bar{X}^2} - \frac{S_{xy}}{\bar{X}\bar{Y}} \right) \frac{1 - (1/N)}{n}$$

donde,

$$x_k \begin{cases} 1 & \text{si el estudiante es un hombre} \\ 0 & \text{en otro caso,} \end{cases}$$

y,

$$y_k \begin{cases} 1 & \text{si el estudiante es una mujer} \\ 0 & \text{en otro caso,} \end{cases}$$

para todo  $k \in U$ . Por lo que, se tiene:

$$S_x^2 = \frac{N}{N-1} P_{.1}(1 - P_{.1})$$

y,

$$S_{xy} = \frac{N}{N-1} P_{11}(1 - P_{.1})$$

En donde,  $P$  hace referencia a los valores poblacionales de las proporciones. Luego, el sesgo  $B(r) \approx 0$ .

Debido a que el sesgo es nulo y  $n$  es grande, se tiene que el error cuadrático medio es aproximadamente igual que la varianza, por lo tiene la forma:

$$\widehat{ECM}(r_H) = \frac{1-f}{n\hat{P}_{.1}} \frac{N}{N-1} \frac{\hat{P}_{11}}{\hat{P}_{.1}} \left( 1 - \frac{\hat{P}_{11}}{\hat{P}_{.1}} \right)$$

así,

$$\widehat{ECM}(r_H) = 0,00379041 \quad \widehat{ECM}(r_M) = 0,0049432148$$

Finalmente, los intervalos de confianza asociados estarán dados por:

$$IC(r_H) = [0,515; 0,757] \quad IC(r_M) = [0,418; 0,694]$$

Debido a que los intervalos se intersección considerablemente, no es posible aseverar que las razones son significativamente diferentes.