Métodos Estadísticos
Tarea 3

Rubén Pérez Palacios Lic. Computación Matemática
Profesor: Dr. Rogelio Ramos Quiroga

3 de marzo de 2022

De acuerdo a la Ley Hardy-Weinberg, los genotipos AA, Aa y aa, deben aparecer en la pobalción con frecuenccis relativas θ^2, $2\theta(1-\theta)$ y $(1-\theta)^2$, respectivamente.

1. En un experimento, para estimar la frecuencia, θ, del gen, se van a elegir al azar, n individuos de la población y se registrarán las frecuencias Y_1, Y_2 y Y_3 de los tres genotipos. Encuentra la función de información esperada de θ.

   Solución 1 – *Notemos que nuestro problema se distribuye multinomial con parametros n, $\left(\theta^2, 2\theta(1-\theta), (1-\theta)^2\right)$. Ahora tenemos que $Y_1, Y_2, Y_3 \sim Multnom(n, \vec{p})$ por lo que la función de verosimilitud de θ es*

$$L(\theta) = C\left(\theta^2\right)^{y_1}(2\theta(1-\theta))^{Y_2}\left((1-\theta)^2\right)^{Y_3},$$

   por lo que su función de logverosimilitud es

$$l(\theta) = C + (2y_1 + y_2)\log(\theta) + (y_2 + 2y_3)\log 1 - \theta,$$

   cuya función score es

$$S(\theta) = \frac{2y_1 + y_2}{\theta} + \frac{y_2 + 2y_3}{1-\theta},$$

   de la cual obtenemos la función de información observada de fisher

$$I(\theta) = \frac{2y_1 + y_2}{\theta^2} + \frac{y_2 + 2y_3}{(1-\theta)^2},$$

   por último obtenemos la función de información de fisher

$$\mathcal{I}(\theta) = \frac{2n\theta^2 + 2n\theta(1-\theta)}{\theta^2} + \frac{2n\theta(1-\theta) + 2n(1-\theta)^2}{(1-\theta)^2} = \frac{2n}{\theta(1-\theta)}.$$

2. Suponga que es muy difícil y caro, distinguir entre los genotipos Aa y aa. Ademáas, suponga que se pueden examinar hasta el triple de individuos si sólo se pide identificar a los genotipos AA. Encuentre la función de información esperada de θ si $3n$ individuos van a ser clasificados como AA o $no - AA$.

   Solución 2 – *Notemos que nuestro problema X se distribuye binomial con parametros $3n$, θ^2. Ahora tenemos que $X \sim Multnom(n, \vec{p})$ por lo que la función de verosimilitud de θ es*

$$L(\theta) = C\left(\theta^2\right)^x \left(1-\theta^2\right)^{3n-x},$$

   por lo que su función de logverosimilitud es

$$l(\theta) = C + 2x\log(\theta) + (3n-x)\log 1 - \theta^2,$$

   cuya función score es

$$S(\theta) = \frac{2x}{\theta} + \frac{2\theta(3n-x)}{1-\theta^2},$$

   de la cual obtenemos la función de información observada de fisher

$$I(\theta) = \frac{2x}{\theta^2} + \frac{2(3n-x)\left(1+\theta^2\right)}{\left((1-\theta)^2\right)^2},$$

   por último obtenemos la función de información de fisher

$$\mathcal{I}(\theta) = \frac{6n\theta^2}{\theta^2} + \frac{2\left(3n-2n\theta^2\right)\left(1+\theta^2\right)}{(1-\theta)^2} = \frac{12n}{(1-\theta^2)}.$$

3. ¿Bajo que circunstancias recomendaría hacer el experimento como en (2.), en vez de como en (1.).

   Solución 3 – *La eficiencia relariva esperada del experimento 2 contra el 1 es*

$$e(\theta) = \frac{\frac{12n}{(1-\theta^2)}}{\frac{2n}{\theta(1-\theta)}} = \frac{6\theta}{(1+\theta)}.$$

   Por lo que

$$e(\theta) \geq 1 \Leftrightarrow \theta \geq \frac{1}{5},$$

   por lo tanto concluimos que usar el experimento 2 cuando $\theta \geq \frac{1}{5}$ y el 1 cuando no.