

[40] Un ingeniero ambiental desea estimar el número total de árboles en un determinado condado que han sido afectados por una plaga y cuál es el nivel de esta infección. Hay 15 zonas forestales bien definidas en el condado, las cuales están divididas en parcelas de aproximadamente el mismo tamaño. Cuatro equipos están disponibles para el estudio, el cual deberá completarse en un día. Con este propósito, se diseñó un muestreo aleatorio por conglomerados bietápico, en el que se seleccionaron al azar y sin reemplazo 4 zonas y 6 parcelas. Los datos obtenidos se presentan en la siguiente tabla:

Zona	Número de parcelas	Número de árboles infectados (con infección avanzada)
1	12	15(5), 14(2), 21(8), 18(3), 9(1), 10(0)
2	16	4(0), 7(2), 10(1), 9(1), 8(3), 5(0)
3	14	10(3), 11(2), 14(2), 10(1), 9(0), 15(4)
4	21	6(2), 3(1), 4(1), 1(0), 2(0), 5(1)

- (a) (20 puntos) Estime la proporción de árboles en el condado que tienen una infección avanzada y confeccione un intervalo del 95 % de confianza para esta cantidad. Interprete sus resultados.
- (b) (10 puntos) Calcule los pesos muestrales para cada una de las observaciones pertenecientes a la muestra, y luego utilice estas cantidades para estimar el total poblacional de árboles infectados.
- (c) (10 puntos) ¿Es posible aseverar que la proporción de árboles con infección avanzada es superior a 18 %? Utilice un $\alpha = 0,1$

Solución: Para obtener la estimación insesgada de la proporción poblacional, utilizamos las fórmulas definidas para la media.

$$\hat{y}_r = \frac{\sum_{i \in \mathcal{S}} \hat{t}_i}{\sum_{i \in \mathcal{S}} M_i} = \frac{\sum_{i \in \mathcal{S}} M_i \bar{y}_i}{\sum_{i \in \mathcal{S}} M_i}$$

en donde \bar{y}_i está dado por la media de las proporciones de árboles con infección avanzada en cada una de las parcelas y M_i son el número de parcelas por zona, se tiene que:

$$\hat{y}_r = 0,1716853$$

Luego, para confeccionar el intervalo de confianza debemos obtener una estimación para la varianza de la estimador de la proporción, usando:

$$\widehat{\mathbb{V}}[\hat{y}_r] = \frac{1}{M^2} \left(1 - \frac{n}{N}\right) \frac{s_r^2}{n} + \frac{1}{nNM^2} \sum_{i \in \mathcal{S}} M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{s_i^2}{m_i}$$

en donde cada $m_i = 6$, y los términos s_i^2 y s_r^2 están dados por:

$$s_i^2 = \frac{1}{m_i - 1} \sum_{j \in S_i} (y_{ij} - \bar{y}_i)^2$$

y,

$$s_r^2 = \frac{1}{n - 1} \sum_{i \in \mathcal{S}} (M_i \bar{y}_i - M_i \hat{y}_r)^2$$

,respectivamente. Así,

$$\sqrt{\widehat{\mathbb{V}}[\hat{y}_r]} = 0,0090104$$

Y en consecuencia, el intervalo de confianza pedido estará definido por:

$$[0,1716853 \pm 1,96 * 0,0090104] = [0,1540249; 0,1893457]$$

Para calcular los pesos muestrales, basta utilizar:

$$w_{ij} = \frac{NM_i}{nm_i} = [7,5 \quad 10 \quad 8,75 \quad 13,125]$$

Luego, utilizando:

$$\hat{t}_{unb} = \sum_{i \in S} \sum_{j \in S_j} w_{ij} y_{ij} = 40,56065$$

en donde los y_{ij} son la proporción de árboles con infección avanzada en cada parcela perteneciente a la muestra.

Para responder el tercer ítem, debemos plantear un test de hipótesis para la proporción poblacional:

$$H_0 : p \leq 0,18 \quad H_1 : p > 0,18$$

El estadístico de prueba bajo H_0 estará dado por:

$$E = \frac{0,1716853 - 0,18}{\sqrt{\frac{0,18 * (1 - 0,18)}{24}}} = -0,1060252$$

Luego, al compararlo con el cuantil respectivo $Z_{1-\alpha}$, se concluye que no existe evidencia estadística suficiente para aseverar que la proporción de árboles con infección avanzada es superior a 18 %.