P. Sea $(a_1, a_2, \dots, a_N)$ una población finita de N objetos e y una variable real, denotamos $y_i = y(a_i)$ el valor que toma y en a_i (Por ejemplo, $y_i =$ el numero de hermanos que la persona i tiene). Sea una muestra M de n objetos de la población obtenidos aleatoriamente con equiprobabilidad y sin reemplazo.

Llamemos $x_i = \begin{cases} 1 & a_i \in M \\ 0 & a_i \notin M \end{cases}$.

e) Calcula la varianza de la media muestral de y en función de los parametros de la población

$$\mu = \sum_{i=1}^N y_i \text{ y } \sigma^2 = \sum_{i=1}^N (y_i - \mu)^2.$$

Sol. Tenemos de las partes anteriores $\bar{y} = \frac{1}{n} \sum_{i=1}^N x_i \cdot y_i$. Luego

$$\mathbb{V}(\bar{y}) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^N x_i \cdot y_i\right) = \frac{1}{n^2} \left(\sum_{i=1}^N y_i^2 \mathbb{V}(x_i) + \sum_{i=1}^N \sum_{j=1, j \neq i}^N y_i y_j \text{cov}(x_i, x_j) \right)$$

Calculemos la covarianza:

$$\text{cov}(x_i, x_j) := \mathbb{E}((x_i - \mathbb{E}(x_i))(x_j - \mathbb{E}(x_j))) = \mathbb{E}(x_i x_j) - \mathbb{E}(x_i)\mathbb{E}(x_j)$$

Usando la definición de probabilidad condicional tenemos:

$$\begin{aligned} \mathbb{E}(x_i x_j) &= 1 \cdot \mathbb{P}(x_i x_j = 1) + 0 \cdot \mathbb{P}(x_i x_j = 0) = \mathbb{P}(x_i = 1, x_j = 1) = \mathbb{P}(a_i, a_j \in M) \\ &= \mathbb{P}(a_i \in M / a_j \in M) \cdot \mathbb{P}(a_j \in M) = \frac{n-1}{N-1} \cdot \frac{n}{N} \end{aligned}$$

Reemplazando queda

$$\text{cov}(x_i, x_j) = \frac{n-1}{N-1} \cdot \frac{n}{N} - \left(\frac{n}{N}\right)^2 = -\frac{n(N-n)}{N^2(N-1)} \text{ si } i \neq j$$

Y así,

$$\mathbb{V}(\bar{y}) = \frac{1}{n^2} \left(\frac{n(N-n)}{N^2} \sum_{i=1}^N y_i^2 - \frac{n(N-n)}{N^2(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N y_i y_j \right)$$

Podemos trabajar un poco más la expresión, considerando que:

$$\sum_{j=1, j \neq i}^N y_i y_j = \sum_{i=1}^N \left(y_i \sum_{j=1}^N (y_j) - y_i^2 \right) = \sum_{i=1}^N y_i (N\mu - y_i) = N^2 \mu^2 - \sum_{i=1}^N y_i^2$$

$$\begin{aligned} \Rightarrow \mathbb{V}(\bar{y}) &= \frac{1}{n^2} \left(\frac{n(N-n)}{N^2} \frac{n(N-n)}{N^2(N-1)} \sum_{i=1}^N y_i^2 - \frac{1}{n^2} \left(\frac{n(N-n)}{N^2(N-1)} N^2 \mu^2 \right) \right) \\ &= \left(\frac{1}{n} \frac{(N-n)(N-1) + (N-n)}{N^2(N-1)} \right) \sum_{i=1}^N y_i^2 - \frac{N-n}{n(N-1)} \mu^2 \\ &= \frac{N-n}{n(N-1)} \left(\frac{1}{N} \sum_{i=1}^N y_i^2 - \mu^2 \right) = \frac{N-n}{n(N-1)} \sigma^2 \end{aligned}$$

El último paso lo justificamos en la primera pregunta (pueden considerarlo una propiedad conocida).