[15] Bajo un muestreo aleatorio simple sin reposición. Considere z_i una variable aleatoria Bernoulli, tal que $z_i = 1$ si la unidad i de la población es incluida en la muestra y $z_i = 0$ en caso contrario. Por lo que:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{N} y_i z_i$$

(a) (7 puntos) Calcule $\mathbb{E}(\overline{y})$
Ayuda: Considere

$$\mathbb{V}(\overline{y}) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{N} y_i z_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^2 \mathbb{V}(z_i) + \sum_{i}^{N}\sum_{i\neq j} y_i y_j cov(z_i, z_j)\right]$$

(b) (5 puntos) Calcule $\mathbb{E}(z_i z_j)$ y utilice esto para calcular $cov(z_i, z_j)$

(c) (3 puntos) Finalmente, considerando la siguiente identidad:

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

Muestre que $\mathbb{V}(\overline{y}) = \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n}$

Solución: Ya que z_i es una variable aleatoria Bernoulli, su valor esperado es $\mathbb{P}(z_i = 1) = n/N$. Así,

$$\mathbb{E}(\overline{y}) = \frac{1}{n}\sum_{i=1}^{N} y_i \mathbb{E}(z_i) = \frac{1}{n}\sum_{i=1}^{N} y_i \frac{n}{N} = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu$$

De igual manera, como z_i es una variable aleatoria Bernoulli, su varianza está dada por:

$$\mathbb{V}(z_i) = (n/N)(1 - n/N).$$

El número de muestras que contienen ambas unidades i y j, cuando $i \neq j$ es $\binom{N-2}{n-2}$, por lo que la probabilidad que ambas unidades estén incluidas es:

$$\frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

Así, el producto $z_i z_j$ será cero excepto en el caso en que ambas unidades i y j estén incluidas en la muestra, por lo que:

$$\mathbb{E}(z_i z_j) = \mathbb{P}(z_i = 1, z_j = 1) = \frac{n(n-1)}{N(N-1)}$$

Luego,

$$cov(z_i, z_j) = \mathbb{E}(z_i z_j) - \mathbb{E}(z_i)\mathbb{E}(z_j) = \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2 = \frac{-n(1-n/N)}{N(N-1)}$$

Considerando la expresión dada para $\mathbb{V}(\overline{y})$, reemplazando los términos $\mathbb{V}(z_i)$ y $cov(z_i, z_j)$ y utilizando la identidad apropiadamente, lo anterior puede ser expresado de la forma:

$$\mathbb{V}(\overline{y}) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\frac{\sum(y_i-\mu)^2}{N-1} = \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n}$$