

TEMA 1. INTRODUCCIÓN A LA INFERENCIA EN POBLACIONES FINITAS: EL MODELO DE POBLACIÓN FIJA Y EL MODELO DE SUPERPOBLACIÓN

1 Introducción

La esencia de las encuestas por muestreo consiste en la selección de una parte de un conjunto de unidades, seguido de la formulación de resultados acerca de la población completa en base a la parte seleccionada. En este proceso el estadístico debe guiarse por el conjunto de principios y procedimientos de la Inferencia Estadística.

Durante mucho tiempo los problemas teóricos de las encuestas por muestreo no preocupaban a los especialistas de la inferencia estadística general, que consideraban dichas encuestas una actividad de tipo práctico. Por otra parte los resultados de la Teoría de la Probabilidad y de la Inferencia Teórica parecían demasiado abstractos a los especialistas en encuestas, inmersos en las dificultades de la puesta en práctica de los muestreos.

Fue a partir de los resultados de *Godambe* (1955) cuando se produjo un cambio de actitud y así la inferencia en muestreo ha tenido un fuerte desarrollo en las últimas décadas.

En el estudio de la inferencia en poblaciones finitas existen dos líneas fundamentales de desarrollo basadas en:

- el punto de vista prevalente en las estadísticas dedicadas al diseño y ejecución de las encuestas por muestreo, de poblaciones finitas, reales y efectivas, en las que el elemento estocástico se debe al mecanismo de aleatorización que determina qué parte de la población se observa, y
- el punto de vista según el cual las muestras proceden de poblaciones que son realizaciones de superpoblaciones o poblaciones infinitas, previamente especificadas, tal y como se venía haciendo en la estadística matemática anterior al muestreo de poblaciones finitas, y en las que el elemento estocástico es introducido por el modelo probabilístico que fijamos a la población.

La divergencia entre dichos puntos de vista se pone de manifiesto especialmente al tratar la aleatorización previa a la selección de la muestra.

De acuerdo con estos puntos de vista, podemos considerar en relación con las encuestas por muestreo, dos tipos de inferencia:

- la inferencia en poblaciones fijas, y
- la inferencia en superpoblaciones o modelos.

El modelo de población fija es el clásico que se ha usado en los trabajos de muestreo en poblaciones finitas. Sin embargo hay algunos ejemplos de modelos de superpoblación que datan de la primera mitad del siglo, como *Cochran* (1939, 1946) o *Deming* y *Stephan* (1941) y en la actualidad muchas de las aportaciones más importantes a la inferencia en poblaciones finitas utilizan la aproximación por superpoblaciones.

2 El modelo de población fija

La inferencia estadística tradicional y la inferencia en muestreo no son teorías opuestas. Sin embargo, los conceptos básicos como parámetro, muestra, dato, estimador, etcétera, tienen un significado especial en muestreo.

Comenzaremos definiendo el modelo “básico” de muestreo. Este modelo incorpora conceptos característicos del muestreo en poblaciones finitas, que en algunos casos como son los conceptos de diseño muestral o esquema muestral no tienen una contrapartida directa en teoría estadística tradicional. Otros como la identificabilidad de la unidades, que usualmente no es tenida en cuenta en la inferencia clásica, va a ser responsable de algunos de los resultados inesperados que proporciona la inferencia en poblaciones finitas.

La formulación del modelo básico es debida esencialmente a *Godambe* (1955), con posteriores modificaciones dadas por el propio *Godambe* y otros. Describiremos el modelo tal y como se usa para la aproximación por población fija, para posteriormente extenderlo a la aproximación por superpoblación.

3 Generalidades

3.1 Definiciones preliminares

Sea una *población finita* $\mathcal{U} = \{u_1, \dots, u_N\}$ constituida por N unidades *distintas e identificables*. N diremos que es el *tamaño de la población* y asumiremos en general que es un valor conocido. Los elementos de \mathcal{U} se llaman *unidades de la población* o *unidades de muestreo*.

Por conveniencia, a veces se referirán las unidades sólo por su subíndice, es decir, la unidad u_i diremos que es la unidad i -ésima, y así representaremos la población por $\mathcal{U} = \{1, \dots, N\}$.

La identificabilidad de las unidades significa que existe una correspondencia uno a uno entre las unidades y los índices. Por ejemplo, en el estudio de un conjunto de 150 pueblos, las unidades poblacionales (pueblos) son distintas y realmente identificables. No obstante, no todas las poblaciones finitas naturales admiten la identificabilidad (por ejemplo, la población de pescado en un lago).

Las unidades poblacionales poseen muchas características de interés, algunas conocidas y otras desconocidas. Estas últimas se llaman *variables de interés* y sus valores (cuantitativos o cualitativos) supondremos que están bien definidos para todas las unidades, y aunque son desconocidos a priori, pueden obtenerse mediante algún procedimiento de investigación (encuesta, procedimiento de medida, etc.).

El propósito fundamental del muestreo es hacer inferencia acerca de alguna función desconocida (llamada *parámetro*) de las variables de interés, basada en los valores que toman las variables en las unidades que constituyen la muestra de la población.

Formalmente, definimos una *muestra ordenada* de una población $\mathcal{U} = \{u_1, \dots, u_N\}$ y la denotaremos por s^* , a un conjunto ordenado de unidades de \mathcal{U} :

$$s^* = (i_1, i_2, \dots, i_{n(s^*)}) \quad \text{con} \quad 1 \leq i_1, i_2, \dots \leq i_{n(s^*)} \leq N \quad (1)$$

donde $n(s^*)$ denota el número de unidades de la muestra que llamaremos *tamaño muestral*.

Llamaremos *tamaño muestral efectivo* y lo denotaremos por $\nu(s^*)$ al número de unidades distintas que componen la muestra.

En general las unidades se mostrarán en una secuencia que corresponde al orden en que han sido seleccionadas e incluso cada unidad con determinada multiplicidad.

Dos muestras s_1^* y s_2^* diremos que son *idénticas* si los vectores son iguales. Dos muestras s_1^* y s_2^* diremos que son equivalentes, $s_1^* \approx s_2^*$, si están compuestas por las mismas unidades. Así, cualquier muestra de tamaño n tiene $n!$ formas diferentes, correspondientes a las posibles permutaciones de sus elementos. Si dos muestras no son equivalentes diremos que son *distintas*.

Ejemplo 3.1 Sea $\mathcal{U} = \{1, 2, 3, 4, 5\}$. Dos muestras particulares serían $s_1^* = (1, 2, 4)$ y $s_2^* = (5, 1)$ con tamaños $n(s_1^*) = 3$ y $n(s_2^*) = 2$, y tamaños efectivos $\nu(s_1^*) = 3$ y $\nu(s_2^*) = 2$. La muestra $s_1^* = (1, 2, 4)$ tiene otras formas equivalentes: $(1, 4, 2)$, $(2, 1, 4)$, $(2, 4, 1)$, $(4, 1, 2)$ y $(4, 2, 1)$, mientras que la muestra $s_2^* = (1, 5)$ tiene por equivalentes, por ejemplo, a $(5, 1)$, $(1, 1, 5)$ y $(1, 5, 5)$, todas de tamaño efectivo 2.

Al conjunto de todas las muestras ordenadas sobre U lo denotaremos por \mathcal{S}^*

Normalmente se está interesado sólo en las unidades que están incluidas en la muestra s^* , independientemente del orden y la multiplicidad con que aparezcan en la secuencia de observación, y se define un conjunto s asociado a la muestra s^* , $s = \{k, k \in s^*\}$. El cardinal de este conjunto es el tamaño muestral efectivo, $\nu(s^*)$. Así, un conjunto no vacío tal que $s \subset \mathcal{U}$ se llama *muestra no ordenada*.

En una población de tamaño N habrá $\binom{N}{n}$ muestras distintas de tamaño n , y al conjunto de todas las muestras distintas lo denotaremos por \mathcal{S} y lo llamaremos *espacio muestral universal*.

El conjunto s se obtiene como una reducción de una secuencia s^* a través de la *función de reducción*, r , definida sobre el conjunto de todas las secuencias, \mathcal{S}^* , de la forma:

$$r : \mathcal{S}^* \longrightarrow \mathcal{S} \quad \forall s^* \in \mathcal{S}^* \exists s \mid r(s^*) = s. \quad (2)$$

Ejemplo 3.2 La población $\mathcal{U} = \{1, 2, 3\}$ tiene siete muestras no ordenadas distintas: $s_1 = \{1\}$, $s_2 = \{2\}$, $s_3 = \{3\}$, $s_4 = \{1, 2\}$, $s_5 = \{1, 3\}$, $s_6 = \{2, 3\}$ y $s_7 = \{1, 2, 3\}$. El espacio muestral universal

será $\mathcal{S} = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$. Habría $3! = 6$ muestras ordenadas de tamaño efectivo 3: $s_{11}^* = (1, 2, 3)$, $s_{12}^* = (2, 3, 1)$, $s_{13}^* = (3, 1, 2)$, $s_{14}^* = (1, 3, 2)$, $s_{15}^* = (2, 1, 3)$ y $s_{16}^* = (3, 2, 1)$, verificándose $r(s_{11}^*) = r(s_{12}^*) = r(s_{13}^*) = r(s_{14}^*) = r(s_{15}^*) = r(s_{16}^*) = s_7$. Sobre la secuencia $s^* = (1, 1, 2)$ la función de reducción daría $r(s^*) = r((1, 1, 2)) = s_4$.

3.2 El diseño muestral

El muestreo en poblaciones finitas se basa normalmente en seleccionar una muestra aleatoriamente para obtener información de sus unidades. El mecanismo aleatorio para la selección de una de estas muestras se formaliza con el concepto de diseño muestral.

Definición 3.3 *Un diseño muestral no ordenado d en una población finita \mathcal{U} , es un par $d = (S_d, P_d)$, donde S_d es un subconjunto del espacio muestral universal \mathcal{S} y P_d es una función de probabilidad definida en S_d tal que:*

- i) $P_d(s) > 0 \quad \forall s \in S_d$, y
- ii) $\forall u \in \mathcal{U} \exists s \in S_d \mid s \ni u$.

En el caso de que la condición ii) no se verifique el diseño muestral estará basado sólo en una subpoblación de \mathcal{U} .

Ejemplo 3.4 *Dada la población $\mathcal{U} = \{1, 2, 3\}$, definimos $d_1 = (S_1, P_1)$ con $S_1 = \{(1), (2), (1, 3)\}$ y $P_1((1)) = \frac{1}{3}$, $P_1((2)) = a$ y $P_1((1, 3)) = \frac{2}{3} - a$, $0 < a < \frac{2}{3}$, con lo que d_1 es un diseño muestral en \mathcal{U} .*

Definimos $d_2 = (S_2, P_2)$ con $S_2 = \{(1, 2), (1, 3), (2, 3)\}$ y $P_2((1, 2)) = P_2((1, 3)) = \frac{1}{4}$ y $P_2((2, 3)) = \frac{1}{2}$, que también es un diseño muestral en \mathcal{U} .

Otro diseño muestral en \mathcal{U} es $d_3 = (S_3, P_3)$ con $S_3 = S_2$ y $P_3((1, 2)) = P_3((1, 3)) = P_3((2, 3)) = \frac{1}{3}$, conocido con el nombre de muestreo aleatorio simple de tamaño 2.

Definición 3.5 *Un diseño muestral ordenado d^* en una población finita \mathcal{U} , es un par $d^* = (S_d^*, P_d^*)$, donde S_d^* es un subconjunto del espacio muestral \mathcal{S}^* y P_d^* es una función de probabilidad definida en S_d^* tal que:*

- i) $P_d^*(s^*) > 0 \quad \forall s^* \in S_d^*$, y

ii) $\forall u \in \mathcal{U} \exists s^* \in S_d^* \mid s^* \ni u$.

Ejemplo 3.6 Definimos $d_4^* = (S_4^*, P_4^*)$ con

$S_4^* = \{(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)\}$ y $P_4((1, 2)) = q_1$, $P_4((2, 1)) = \frac{1}{4} - q_1$, $P_4((1, 3)) = q_2$, $P_4((3, 1)) = \frac{1}{4} - q_2$, $P_4((2, 3)) = q_3$, $P_4((3, 2)) = \frac{1}{2} - q_3$, con $0 < q_1, q_2 < \frac{1}{4}$ y $0 < q_3 < \frac{1}{2}$, que es un diseño muestral ordenado en \mathcal{U} .

Definición 3.7 Se llama versión simetrizada o reducida de un diseño muestral ordenado $d^* = (S_d^*, P_d^*)$ al diseño que se obtiene definiendo $S_d = r(S_d^*)$ y

$$P_d(s) = \sum_{s^* \approx s} P_d(s^*) \quad \forall s \in S_d.$$

Las muestras s^* son muestras ordenadas y las muestras s muestras no ordenadas.

Ejemplo 3.8 $(1, 2), (1, 3), (2, 3)$ son muestras no ordenadas y $(1, 2), (2, 1), (1, 3), (3, 1), (1, 4), (4, 1)$ son muestras ordenadas.

Definición 3.9 Se llama soporte de un diseño muestral $d = (S_d, P_d)$ en una población \mathcal{U} a S_d y el tamaño del soporte muestral es el cardinal de S_d .

Definición 3.10 Un diseño muestral $d = (S_d, P_d)$ en una población \mathcal{U} se dice que es uniforme si P_d es uniforme en S_d .

Definición 3.11 Un diseño muestral $d = (S_d, P_d)$ en una población \mathcal{U} se dice que es de tamaño fijo n , si todas las muestras tienen el mismo tamaño, $n(s) = n$.

Ejemplo 3.12 En los ejemplos anteriores (1.3) y (1.4), el diseño d_3 es uniforme, mientras que d_1 , d_2 y d_4 no lo son. d_2 , d_3 y d_4 son diseños de tamaño fijo y d_1 es de tamaño variable. El diseño d_1 es una versión reducida del diseño d_4 puesto que $P_1((1, 2)) = P_4((1, 2)) + P_4((2, 1))$, $P_1((1, 3)) = P_4((1, 3)) + P_4((3, 1))$ y $P_1((2, 3)) = P_4((2, 3)) + P_4((3, 2))$.

3.3 Algunos tipos de diseños muestrales no ordenados

3.3.1 Diseño muestral universal uniforme

En este diseño $S_d = \mathcal{S}$, es decir, todas las posibles muestras que pueden ser elegidas, y $P_d = \frac{1}{2^N}$, $\forall s \in S_d$. Este diseño no es realista pues, por ejemplo, da igual importancia a una muestra de tamaño 1 que a una muestra que contenga a toda la población.

3.3.2 Diseño muestral de Bernouilli

En él $S_d = \mathcal{S}$ y $P_d(s) = p^{n(s)}(1-p)^{N-n(s)}$ con $0 < p < 1$ un parámetro. Por tanto, las muestras de tamaño dado $n(s) = k$ acumulan una probabilidad de $\binom{N}{k} p^k (1-p)^{N-k}$ (distribución binomial), y por tanto, las muestras con pocos o muchos elementos son menos probables que las muestras con un tamaño próximo al tamaño promedio Np . La obtención de una muestra según este diseño puede hacerse generando N números aleatorios entre 0 y 1, $\varepsilon_1, \dots, \varepsilon_n$, de modo independiente y seleccionando la unidad u_i si $\varepsilon_i < p$.

3.3.3 Diseño muestral de tamaño fijo uniforme

Es conocido como muestreo aleatorio simple, denotado por $MAS(N, n)$, y en él S_d es el conjunto de todas las posibles muestras distintas s de tamaño fijo $n(s) = n$ y $P_d(s) = \frac{1}{\binom{N}{n}}$, $\forall s \in S_d$.

3.3.4 Diseño unicluster

Es un diseño en el cual todas las parejas de muestras que pertenecen al soporte S_d son disjuntas o equivalentes.

3.3.5 Diseño muestral cluster uniforme

La población se divide en k partes disjuntas de tamaño n , siendo $N = kn$. Cada muestra de S_d es una de estas k partes y la función de probabilidad es la uniforme en S_d , es decir, denotando $S_d = \{s_1, \dots, s_k\}$ con $s_i \cap s_j = \emptyset$ si $i \neq j$ y $\cup_i^k s_i = \mathcal{U}$, $P_d(s_i) = \frac{1}{k}$, $\forall i$.

3.3.6 Muestreo sistemático

Un ejemplo conocido del diseño anterior es el muestreo sistemático uniforme de paso k , $MS(N, k)$. Tiene por espacio muestral muestras sistemáticas, es decir, las unidades muestreadas se escogen de la población mediante la elección de una unidad de partida y las demás por una regla sistemática de elección, con $N/k = n$ entero. El espacio muestral está dado por $S_d = \{s_1, \dots, s_k\}$, siendo $s_t = \{t, t+k, \dots, t+(n-1)k\}$, $t = 1, \dots, k$ y la distribución de probabilidad es la uniforme en S_d , es decir, $p(s_t) = \frac{1}{k}$, $\forall s_t \in S_d$.

3.3.7 Diseño muestral estratificado

Una población \mathcal{U} se dice estratificada en L estratos U_1, \dots, U_L si estos forman una partición de la misma. En cada estrato U_h se define un diseño muestral $d_h = (S_h, p_h)$ independientemente de los demás estratos, obteniéndose en él una muestra s_h . El diseño muestral resultante en \mathcal{U} viene dado por $S_d = S_1 \times S_2 \times \dots \times S_L$ y la probabilidad de una muestra $s = (s_1, \dots, s_L)$ es $p(s) = p_1(s_1)p_2(s_2) \dots p_L(s_L)$.

3.3.8 Diseño muestral de conglomerados

La población $\mathcal{U} = \{1, 2, \dots, N\}$ está dividida en conglomerados C_1, C_2, \dots, C_M , cada uno de ellos constituido por unidades finales de \mathcal{U} , $C_i = \{i_1, \dots, i_{N_i}\}$, de forma que constituyen una partición del espacio. De esta población de conglomerados $U_c = \{C_1, \dots, C_M\}$ se extrae una muestra de conglomerados $s_c = \{C_{j_1}, \dots, C_{j_g}\}$ y dado un conglomerado $C_j \in s_c$, se observan todas las unidades que lo componen. Así, la muestra final está constituida por todas las unidades finales que pertenecen a los conglomerados de la muestra s_c .

3.4 Probabilidades de inclusión y matriz de diseño

Como se ha visto anteriormente (definición 1.1), cada unidad de la población debe estar en alguna muestra, pero hay unidades que pueden estar en más de una. Así, cada unidad puede tener una probabilidad distinta de pertenecer a la muestra elegida. Esta probabilidad se cuantifica mediante las probabilidades de inclusión.

Definición 3.13 Para cada unidad i de la población \mathcal{U} , se define la probabilidad de inclusión de primer orden bajo el diseño muestral $d = (S_d, P_d)$ como:

$$\pi_i(d) = \sum_{s \ni i} P_d(s) \quad \forall i. \quad (3)$$

La suma anterior se extiende a todas las muestras del diseño d que contengan a la unidad i -ésima.

Definición 3.14 Para cada dos unidades i, j de la población \mathcal{U} , se define la probabilidad de inclusión de segundo orden bajo el diseño muestral $d = (S_d, P_d)$ como:

$$\pi_{ij}(d) = \sum_{s \ni i, j} P_d(s) \quad \forall i, j. \quad (4)$$

La suma anterior se extiende a todas las muestras del diseño d que contengan simultáneamente a las unidades i -ésima y j -ésima.

Definición 3.15 Un diseño muestral $d = (S_d, P_d)$ en una población \mathcal{U} se llama probabilístico cuando $\pi_i(d) > 0, \forall i \in \mathcal{U}$.

Definición 3.16 Un diseño muestral $d = (S_d, P_d)$ en una población \mathcal{U} se llama cuantificable cuando $\pi_{ij}(d) > 0 \forall i, j \in \mathcal{U}$.

Por simplicidad y cuando no de lugar a confusión se denotarán las probabilidades de inclusión de primer y segundo orden por π_i y π_{ij} , respectivamente.

Definición 3.17 Se llama matriz de diseño del diseño muestral d de la población \mathcal{U} a la matriz $N \times N$ simétrica $\Pi(d) = (\pi_{ij})_{1 \leq i, j \leq N}$.

Ejemplo 3.18 Las probabilidades de inclusión del diseño d_1 introducido en el ejemplo (1.3) son: $\pi_1 = P_1((1, 1)) + P_1((1, 3)) = \frac{1}{3} + \frac{2}{3} - a = 1 - a$, $\pi_2 = P_1((2)) = a$, $\pi_3 = P_1((1, 3)) = \frac{2}{3} - a$, $\pi_{12} = 0$, $\pi_{13} = \frac{2}{3} - a$ y $\pi_{23} = 0$. La matriz del diseño d_1 sería:

$$\Pi(d_1) = \begin{pmatrix} 1 - a & 0 & \frac{2}{3} - a \\ 0 & a & 0 \\ \frac{2}{3} & 0 & \frac{2}{3} - a \end{pmatrix}$$

Las probabilidades de inclusión del diseño d_2 son: $\pi_1 = P_2((1, 2)) + P_2((1, 2)) = \frac{1}{2}$, $\pi_2 = P_2((1, 2)) + P_2((2, 3)) = \frac{3}{4}$, $\pi_3 = P_2((1, 3)) + P_2((2, 3)) = \frac{3}{4}$, $\pi_{12} = P_2((1, 2)) = \frac{1}{4}$, $\pi_{13} = P_2((1, 3)) = \frac{1}{4}$ y $\pi_{23} = P_2((2, 3)) = \frac{1}{2}$. La matriz del diseño d_2 sería:

$$\Pi(d_1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

De la definición de probabilidad de inclusión de primer y segundo orden se puede deducir que, $\forall i, j$, $\max\{0, \pi_i + \pi_j - 1\} \leq \pi_{ij} \leq \min\{\pi_i, \pi_j\}$. Dado un diseño muestral d , siempre se puede encontrar su matriz de diseño asociada Π que verifica la condición anterior.

La importancia de esta matriz radica en que la mayoría de los estimadores y de los errores de muestreo van a poder expresarse en función de los elementos de esta matriz. Interesará entonces comprobar algunas propiedades que relacionan las probabilidades de inclusión de primer y segundo orden entre sí y las relaciones de éstas con el tamaño muestral. También es usual expresar estimadores y errores de muestreo en función de variables que indiquen la pertenencia o no de cada unidad a la muestra. Estas variables se definen a continuación y posteriormente se enuncian algunas de las propiedades que las relacionan con las probabilidades de inclusión de primer y segundo orden.

Definición 3.19 Para cada unidad i de una población finita \mathcal{U} sobre la que hay definido un diseño muestral $d = (S_d, P_d)$ se define su variable indicadora de la forma:

$$I_i(s) = \begin{cases} 1 & \text{si } i \in s \\ 0 & \text{si } i \notin s \end{cases} \quad \forall s \in S_d.$$

Proposición 3.20 Sea d un diseño en una población \mathcal{U} , π_i y π_{ij} las probabilidades de inclusión de primer y segundo orden de las unidades e I_i la variable indicadora de cada unidad de la población. Se verifican las siguientes relaciones, para cada $s \in S_d$:

- i) $E(I_i(s)) = \pi_i$, $\forall i$,
- ii) $V(I_i(s)) = \pi_i(1 - \pi_i)$, $\forall i$,

iii) $\Delta_{ij} = \text{cov}(I_i(s), I_j(s)) = \pi_{ij} - \pi_i \pi_j, \forall i \neq j.$

Demostración.–

- i) $E(I_i(s)) = \sum_{i=1}^N I_i(s) P_d(s) = \sum_{s \ni i} P_d(s) = \pi_i, \forall i,$
- ii) $E(I_i^2(s)) = \sum_{i=1}^N I_i^2(s) P_d(s) = \sum_{s \ni i} P_d(s) = \pi_i, \forall i,$ y entonces
 $V(I_i(s)) = E(I_i^2(s)) - (E(I_i(s)))^2 = \pi_i - \pi_i^2 = \pi_i(1 - \pi_i), \forall i,$
- iii) Como $E(I_i(s), I_j(s)) = \sum_{i=1}^N \sum_{j=1}^N I_i(s) I_j(s) P_d(s) = \sum_{s \ni i, j} P_d(s) \pi_{ij},$ entonces

$$\text{cov}(I_i(s), I_j(s)) = E(I_i(s), I_j(s)) - E(I_i(s))E(I_j(s)) = \pi_{ij} - \pi_i \pi_j, \forall i \neq j.$$

Proposición 3.21 *Sea d un diseño de tamaño no fijo y π_i y π_{ij} las probabilidades de inclusión de primer y segundo orden. Sean $E(n(s))$ y $V(n(s)) = E(n(s)^2) - E(n(s))^2$ la esperanza y varianza del tamaño muestral. Se verifican las siguientes relaciones:*

- i) $\sum_i^N \pi_i = E(n(s)),$
- ii) $\sum_{i=1}^N \sum_{j=1}^N \pi_{ij} = E(n(s)^2),$
- iii) $\sum_{i=1}^N \sum_{j(\neq i)=1}^N \pi_{ij} = E(n(s)(n(s) - 1)),$
- iv) para cada i fijo, $\sum_{j(\neq i)=1}^N \pi_{ij} = E(I_i(s)(n(s) - 1)).$

De esta forma, $V(n(s)) = \sum_{i=1}^N \sum_{j=1}^N \pi_{ij} - (\sum_{i=1}^N \pi_i)^2.$

Demostración.–

- i) $\sum_i^N \pi_i = \sum_i^N E(I_i(s)) = E(\sum_i^N I_i(s)) = E(n(s)),$
- ii) $\sum_{i=1}^N \sum_{j=1}^N \pi_{ij} = \sum_{i=1}^N \sum_{j=1}^N E(I_i(s) I_j(s)) =$
 $E(\sum_{i=1}^N \sum_{j=1}^N I_i(s) I_j(s)) = E((\sum_{i=1}^N I_i(s))(\sum_{j=1}^N I_j(s))) = E(n(s)^2),$
- iii) $\sum_{i=1}^N \sum_{j(\neq i)=1}^N \pi_{ij} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N E(I_i(s) I_j(s)) = E(\sum_{i=1}^N \sum_{j(\neq i)=1}^N I_i(s) I_j(s)) =$
 $E((\sum_{i=1}^N I_i(s))(\sum_{j(\neq i)=1}^N I_j(s))) = E(n(s)(n(s) - 1)),$
- iv) para cada i fijo, $\sum_{j(\neq i)=1}^N \pi_{ij} = \sum_{j(\neq i)=1}^N E(I_i(s) I_j(s)) = E(\sum_{j(\neq i)=1}^N I_i(s) I_j(s)) =$
 $E(I_i(s) \sum_{j(\neq i)=1}^N I_j(s)) = E(I_i(s)(n(s) - 1)).$

Corolario 3.22 *Sea d un diseño de tamaño fijo y π_i y π_{ij} las probabilidades de inclusión de primer y segundo orden. Se verifican las siguientes relaciones:*

i) $\sum_i^N \pi_i = n,$

ii) $\sum_{i=1}^N \sum_{j(\neq i)=1}^N \pi_{ij} = n(n-1),$

ii) *para cada i fijo, $i = 1, \dots, N$, $\sum_{j=1(j \neq i)}^N \pi_{ij} = (n-1)\pi_i$*