

TEMA 2. ELEMENTOS DE INFERENCIA

1 Datos y parámetros

Sea s^* una muestra ordenada seleccionada de una población finita \mathcal{U} con un diseño $d^* = (S_d^*, P_d^*)$.

Definición 1.1 *El dato δ^* obtenido de la muestra $s^* \in S_d^*$ viene dado por la secuencia de pares de valores que forman la “etiqueta” (número que identifica la unidad en la población) y la observación de la variable y en esa unidad:*

$$\delta^* = ((i_1, y_1), \dots, (i_{n(s^*)}, y_{n(s^*)}))$$

o escrito en forma más compacta, utilizando la versión reducida de s^ , s :*

$$\delta = ((i, y_i) \mid i \in s, y \in R).$$

El espacio muestral de la variable aleatoria \mathcal{D}^ que toma los valores δ^* lo notaremos por $\Omega^* = \{\delta^* \mid s^* \in S_d^*, \mathbf{y} \in R^N\}$ y el espacio muestral de la variable aleatoria \mathcal{D} que toma los valores δ lo notaremos por $\Omega = \{\delta \mid s \in S_d, \mathbf{y} \in R^N\}$.*

Definición 1.2 *Un dato δ se dice que es consistente con un vector poblacional particular $\mathbf{y}_0 = (y_{01}, \dots, y_{0N})$ si y sólo si $y_i = y_{0i}, \forall i \in s$.*

Así, un dato es consistente con un vector poblacional particular cuando los valores de la variable y para las unidades de la muestra s coinciden con los valores de la variable y para las mismas unidades en el vector poblacional.

Teorema 1.3 *Para un diseño muestral ordenado d^* y para cada vector poblacional $\mathbf{y} \in R^N$, la probabilidad de que la variable aleatoria \mathcal{D}^* tome el valor δ^* viene dada por:*

$$P_{\mathbf{y}}(\delta^*) = \begin{cases} P_d^*(s) & \text{si } \delta^* \text{ es consistente con } \mathbf{y} \\ 0 & \text{en otro caso} \end{cases}$$

Análogamente, si el diseño es no ordenado.

Ejemplo 1.4 Sea $\mathcal{U} = \{1, 2, 3\}$, $\mathbf{y}_0 = (a, b, c)$ y el diseño muestral uniforme de tamaño fijo 2. El dato $\delta_1 = \{(1, a), (2, b)\}$ es consistente con \mathbf{y}_0 y el dato $\delta_2 = \{(1, b), (2, a)\}$ no lo es. En este caso, $P(\mathcal{D} = \delta_1) = P(\delta_1) = P((1, 2)) = \frac{1}{3}$ y $P(\mathcal{D} = \delta_2) = P(\delta_2) = 0$.

Definición 1.5 Un parámetro es una función real de los valores y_i asociados a las unidades u_i o i de la población finita \mathcal{U} , que habitualmente denotaremos por $\theta(\mathbf{y})$, $\mathbf{y} \in R^N$.

A priori, el vector paramétrico $\mathbf{y} = (y_1, \dots, y_N)$ es desconocido, pero asumimos que se puede observar el valor y_i (con o sin error) observando la unidad u_i , $i = 1, \dots, N$. El propósito es hacer inferencia acerca de alguna función $\theta(\mathbf{y})$ desconocida. Las funciones de interés más usuales son el total, $Y = \sum_{i=1}^N y_i$, la media $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$, una proporción, $P = \frac{1}{N} \sum_{i=1}^N A_i$, escrita en función de la variable A que indique la pertenencia o no de cada unidad al grupo, y la varianza, $\sigma^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{Y}^2$.

Definición 1.6 Un parámetro se llama lineal si es de la forma

$$\theta(\mathbf{y}) = a_0 + \sum_{i \in \mathcal{U}} a_i y_i \quad \text{con} \quad a_0, a_1, \dots, a_N \in R.$$

Definición 1.7 Un parámetro se llama lineal y homogéneo si es de la forma

$$\theta(\mathbf{y}) = \sum_{i \in \mathcal{U}} a_i y_i \quad \text{con} \quad a_i \in R, \forall i.$$

Ejemplo 1.8 La media poblacional, $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$, o el total poblacional, $Y = \sum_{i=1}^N y_i$, de una variable y en una población \mathcal{U} son parámetros lineales y homogéneos, con coeficientes $a_i = \frac{1}{N}$ y $a_i = 1$, $\forall i$, respectivamente. La proporción de unidades de la población finita que pertenecen a un determinado grupo o presentan determinada característica, P , también es un parámetro lineal y homogéneo, con coeficientes $a_i = \frac{1}{N}$, $\forall i$. Estos tres parámetros son los parámetros que con mayor frecuencia se está interesado en estimar en el muestreo de poblaciones finitas.

2 Estadísticos y estimadores

Definición 2.1 *Un estadístico es una función real de la muestra del diseño que se considere $d = (S_d, P_d)$ que depende de \mathbf{y} sólo a través de los y_i para los que $i \in s$:*

$$e : S_d \times R^N \longrightarrow R.$$

Definición 2.2 *Un estimador de un parámetro $\theta(\mathbf{y})$ es un estadístico, $e(s, \mathbf{y})$, cuyos valores se emplean para predecir el valor del parámetro de interés $\theta(\mathbf{y})$.*

Por tanto, un estimador depende de \mathbf{y} sólo a través de los y_i para los que $i \in s$, donde s es una muestra del diseño que se considere $d = (S_d, P_d)$ en la población finita \mathcal{U} .

Definición 2.3 *Se llama versión reducida o simetrizada de un estimador $e(s, \mathbf{y})$ de un parámetro $\theta(\mathbf{y})$ al estimador*

$$e' = \sum_{s' \approx s} e(s', \mathbf{y}) P_d(s') / \sum_{s' \approx s} P_d(s'),$$

donde la sumatoria se extiende a todas las muestras s' del diseño muestral d que son equivalentes a s .

Definición 2.4 *Un estimador se llama lineal si es de la forma*

$$e(s, \mathbf{y}) = a_{0s} + \sum_{i \in s} a_{is} y_i$$

donde $a_{0s}, a_{is}, i \in s$ son números reales que no dependen de los valores $\mathbf{y} \in R^N$.

Definición 2.5 *Un estimador se llama lineal y homogéneo si es de la forma*

$$e(s, \mathbf{y}) = \sum_{i \in s} a_{is} y_i$$

donde $a_{is}, i \in s$ son números reales que no dependen de los valores $\mathbf{y} \in R^N$.

Definición 2.6 Un estimador se dice que es independiente de las etiquetas si puede construirse sin el conocimiento de las etiquetas que corresponden al dato δ y en caso contrario se dice que es dependiente de las etiquetas.

Ejemplo 2.7 La media muestral, $\bar{y}_s = \frac{1}{n(s)} \sum_{i=1}^{n(s)} y_i$, es un estimador lineal y homogéneo, con coeficientes $a_{is} = \frac{1}{n(s)}$, $\forall i \in s$, siendo s una muestra del diseño que se considere.

Puesto que sobre el soporte del diseño muestral S_d hay definida una probabilidad P_d , un estimador es una variable aleatoria, cuya función de distribución asociada recibe el nombre de *distribución en el muestreo del estimador*.

Ejemplo 2.8 Sea la población $\mathcal{U} = \{1, 2, 3\}$ y sobre ella el diseño muestral d con soporte $S_d = \{(1, 2), (1, 3), (2, 3)\}$ y probabilidad uniforme, $P_d(s) = \frac{1}{3}$, $\forall s \in S_d$.

La función $e : S_d \times R^3 \longrightarrow R$ dada por

$$e = \begin{cases} \frac{y_3}{2} + \frac{y_2}{2} & \text{si ocurre } s_1 = (1, 2) \\ \frac{y_1}{2} + 2\frac{y_3}{3} & \text{si ocurre } s_2 = (1, 3) \\ \frac{y_2}{2} + 2\frac{y_3}{3} & \text{si ocurre } s_2 = (1, 3) \end{cases}$$

no sería ni siquiera un estadístico pues depende, en el caso de que se seleccione la muestra s_1 de un valor de y , y_3 , para el que $i = 3 \notin s_1$. Redefiniéndolo de forma que tome el valor $\frac{y_1}{2} + \frac{y_2}{2}$ si ocurre $s_1 = (1, 2)$, sería un estadístico lineal y homogéneo para el que $a_{1s_1} = a_{1s_2} = a_{2s_1} = a_{2s_3} = \frac{1}{2}$, $a_{3s_2} = \frac{2}{3}$ y $a_{3s_3} = \frac{1}{3}$. Su distribución en el muestreo vendría dada por:

$$e = \begin{cases} \text{valor} & \text{probabilidad} \\ \frac{y_1}{2} + \frac{y_2}{2} & 1/3 \\ \frac{y_1}{2} + 2\frac{y_3}{3} & 1/3 \\ \frac{y_2}{2} + \frac{y_3}{3} & 1/3 \end{cases}$$

Ejemplo 2.9 Son ejemplos de estimadores los siguientes: $e_1 = \sum_{i \in s} \frac{y_i}{n(s)}$, $e_2 = \sum_{i \in s} y_i$, $e_3 = \min_{i \in s} y_i$,

$e_4 = \max_{i \in s} y_i$, $e_5 = \min_{i \in s} y_i + \max_{i \in s} y_i$, que son variables aleatorias puesto que para cada $s \in S_d$ toman un valor diferente.

Para el diseño muestral $d_2 = (S_2, P_2)$ con $S_2 = \{(1, 2), (1, 3), (2, 3)\}$ con $P_2(1, 2) = P_2(1, 3) = \frac{1}{4}$ y $P_2(2, 3) = \frac{1}{2}$, los estimadores anteriores tienen por distribuciones las siguientes:

$$e_1 = \begin{Bmatrix} 3/2 & 1/4 \\ 4/2 & 1/4 \\ 5/2 & 1/2 \end{Bmatrix}, \quad e_2 = e_5 = \begin{Bmatrix} 3 & 1/4 \\ 4 & 1/4 \\ 5 & 1/2 \end{Bmatrix}, \quad e_3 = \begin{Bmatrix} 1 & 1/2 \\ 2 & 1/2 \end{Bmatrix}, \quad e_4 = \begin{Bmatrix} 2 & 1/2 \\ 3 & 1/2 \end{Bmatrix}$$

2.1 Estimadores usuales

Cuando el parámetro de interés sea la media de una variable y , $\theta(\mathbf{y}) = \sum_{i=1}^N y_i$, (parámetro lineal y homogéneo) en una población finita \mathcal{U} sobre la que haya definido un diseño muestral $d = (S_d, P_d)$, son conocidos diferentes estadísticos que se usan para predecir el parámetro media poblacional, es decir, estimadores de la media poblacional, que después son adaptados para estimar otros parámetros, y que recapitulamos a continuación.

2.1.1 Media muestral

$$\bar{y}_s = \sum_{i \in s} \frac{y_i}{n(s)},$$

estimador lineal homogéneo en el que $a_{is} = \frac{1}{n(s)}$, $\forall i$ y donde s es una muestra de tamaño $n(s)$ de un diseño d . Se trata de un estimador independiente de las etiquetas puesto que su construcción no requiere (o da el mismo resultado) el conocimiento de las secuencia de pares de valores que indique cómo se han obtenido las unidades que componen la muestra obtenida s .

2.1.2 Estimador de razón

Se usa, como veremos más adelante, cuando se dispone de una variable adicional x , para la que es conocido el parámetro media poblacional, $\bar{X} = \sum_{i=1}^N x_i$. En este caso, como en los estimadores de diferencia, regresión y media de razones que definiremos a continuación, se observan en la muestra s las variables x e y , con lo que el dato obtenido de la muestra s sería el conjunto

$$\delta = ((i, y_i, x_i) \mid i \in s, y_i, x_i \in R).$$

Se define como

$$\bar{y}_R = \bar{y}_s \frac{\bar{X}}{\bar{x}_s},$$

donde \bar{x}_s es el estimador media muestral para el parámetro media poblacional de la variable x en la muestra s , y es independiente de las etiquetas.

2.1.3 Estimador media de razones

En este caso, también se dispone de una variable adicional, x , pero no se requiere el conocimiento de ningún parámetro poblacional suyo:

$$\bar{r}_{yx} = \frac{1}{n(s)} \sum_{i \in s} \frac{y_i}{x_i},$$

y el estimador se adapta, como veremos después, para servir como estimador de la media de la variable de interés y . También es independiente de las etiquetas.

2.1.4 Estimador de Horvitz–Thompson

En diseños probabilísticos, es decir, para los que las probabilidades de inclusión de primer orden son estrictamente positivas, $\pi_i > 0$, $i = 1, \dots, N$, Horvitz y Thompson definen el estimador de la media:

$$\bar{y}_{HT} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i},$$

estimador lineal homogéneo con coeficientes $a_{is} = \frac{1}{N\pi_i}$, $\forall i \in s$. Por tanto, se trata de un estimador dependiente de las etiquetas puesto que su construcción conlleva el cálculo de las probabilidades de inclusión de las unidades en la muestra seleccionada s y éstas, a su vez, dependerán de que en el diseño d que se considere influya la secuencia del dato δ .

2.1.5 Estimador de diferencia

Como en el caso del estimador de razón, es un estimador independiente de las etiquetas que se usa con información de una variable auxiliar x con parámetro media conocido:

$$\bar{y}_D = \bar{y}_s + (\bar{X} - \bar{x}_s).$$

2.1.6 Estimador de regresión

Extiende el estimador de diferencia anterior y se usa en el mismo ambiente. Se define como:

$$\bar{y}_D = \bar{y}_s + b(\bar{X} - \bar{x}_s).$$

2.1.7 Estimador diferencia generalizado

Dados e_i , $i = 1 \dots, N$ números reales conocidos, *Basu* (1971) define el estimador

$$\bar{y}_{DG} = \sum_{i \in s} \frac{y_i - e_i}{N\pi_i} + \bar{e},$$

donde $\bar{e} = \sum_{i=1}^N \frac{e_i}{N}$. Este estimador requiere, como el estimador de Horvitz–Thompson, que el diseño sea probabilístico.

3 Características, propiedades y criterios de selección de estadísticos y estimadores

Como en la inferencia estadística clásica, hay propiedades de los estimadores que se utilizan para preferir unos estimadores a otros. Entre ellas se encuentra la propiedad que alude al hecho de que si un estimador es un estadístico que va a ser utilizado para predecir un parámetro, sería de desear que sus valores los tomase centrados en el parámetro que estima. La propiedad de insesgadez se resume en las tres siguientes definiciones.

3.1 Esperanza, varianza y error cuadrático medio

Sea $e(s, \mathbf{y})$ un estimador de un parámetro $\theta(\mathbf{y})$ de una población finita \mathcal{U} sobre la que está definido el diseño muestral $d = (S_d, P_d)$.

Definición 3.1 Se llama esperanza del estimador $e(s, \mathbf{y})$ a

$$E(e(s, \mathbf{y})) = \sum_{s \in S_d} e(s, \mathbf{y}) P_d(s).$$

Definición 3.2 Se llama sesgo del estimador $e(s, \mathbf{y})$ a $\text{sesgo}(e(s, \mathbf{y})) = E(e(s, \mathbf{y})) - \theta(\mathbf{y})$.

Definición 3.3 Se dice que el estimador $e(s, \mathbf{y})$ es insesgado del parámetro $\theta(\mathbf{y})$ cuando $\text{sesgo}(e(s, \mathbf{y})) = 0$, es decir, cuando $E(e(s, \mathbf{y})) = \theta(\mathbf{y}), \forall \mathbf{y} \in R^N$.

Como se comentó anteriormente los parámetros que con más frecuencia se está interesado en estimar son medias, totales y proporciones, y estos son parámetros lineales y homogéneos. Es lógico buscar para un parámetro con estas características un estimador lineal y homogéneo y si una de las condiciones que nos haga preferir unos estimadores a otros va a ser la insesgadez, tiene sentido verificar cuales son las condiciones que tienen que cumplir este tipo de estimadores. Se enuncian en la siguiente proposición.

Proposición 3.4 Un estimador lineal y homogéneo, $e(s, \mathbf{y}) = \sum_{i \in s} a_{is} y_i$ es insesgado del parámetro $\theta(\mathbf{y}) = \sum_{i=1}^N a_i y_i$ si y sólo si $\sum_{s \ni i} a_{is} P_d(s) = a_i, \forall i \in s$.

Definición 3.5 Se llama varianza del estimador $e(s, \mathbf{y})$ a

$$V(e(s, \mathbf{y})) = E(e(s, \mathbf{y}) - E(e(s, \mathbf{y})))^2.$$

Definición 3.6 Se llama error cuadrático medio del estimador $e(s, \mathbf{y})$ a

$$ECM(e(s, \mathbf{y})) = E(e(s, \mathbf{y}) - \theta(\mathbf{y}))^2.$$

Es fácil deducir que $\text{ECM}(e(s, \mathbf{y})) = V(e(s, \mathbf{y})) + \text{sesgo}(e(s, \mathbf{y}))^2$ con lo que en los estimadores insesgados la varianza y el error cuadrático medio coinciden. Entonces, la precisión de un estimador la mediremos con la raíz cuadrada del error cuadrático medio, que se conoce con el nombre de *error de muestreo*, que coincidirá con la *desviación estándar* (raíz cuadrada de la varianza) cuando el estimador sea insesgado.

Hay veces que es útil dar una medida de dispersión adimensional, con el fin de poder comparar precisiones de estimadores de parámetros que vengan medidos en distintas unidades. Daremos la siguiente:

Definición 3.7 *Se llama coeficiente de variación del estimador $e(s, \mathbf{y})$ a*

$$CV(e(s, \mathbf{y})) = \frac{\sqrt{V(e(s, \mathbf{y}))}}{E(e(s, \mathbf{y}))}.$$

3.2 Comparación de estimadores

Dado un diseño muestral $d = (S_d, P_d)$ en una población finita \mathcal{U} y un parámetro $\theta(\mathbf{y})$, uno de los criterios que se emplean para preferir unos estimadores a otros es la insesgadez, si bien que un estimador sea insesgado no significa que sea bueno. Además pueden existir varios estimadores insesgados de un mismo parámetro, lo que hace dar otro criterio para elegir unos estimadores frente a otros. El criterio que se suele seguir es elegir aquél cuya dispersión sea más pequeña o dicho de otra forma cuya precisión sea mayor. Esto motiva las siguientes definiciones.

Definición 3.8 *Un estimador $e_1(s, \mathbf{y})$ se dice que es como mínimo igual de preciso que otro $e_2(s, \mathbf{y})$ cuando*

$$E(e_1(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \leq E(e_2(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \quad \forall \mathbf{y} \in R^N.$$

Cuando en algún $\mathbf{y} \in R^N$ la desigualdad es estricta se dice que e_1 es más preciso que e_2 .

3.2.1 Admisibilidad

Definición 3.9 *Un estimador $e_2(s, \mathbf{y})$ se dice que es inadmisibile para el parámetro $\theta(\mathbf{y})$ cuando existe otro estimador $e_1(s, \mathbf{y})$ y algún $\mathbf{y} \in R^N$ donde*

$$E(e_1(s, \mathbf{y}) - \theta(\mathbf{y}))^2 < E(e_2(s, \mathbf{y}) - \theta(\mathbf{y}))^2.$$

Así, la admisibilidad viene a ser un requisito mínimo que ha de cumplir un estimador de un parámetro.

El teorema que veremos a continuación pone de manifiesto cómo la versión simetrizada, e^* , de cualquier estimador, e , es como mínimo igual de preciso que e , y que permitirá quedarse sólo con las formas reducidas de los diseños muestrales en los problemas de inferencia.

Teorema 3.10 *El estimador e' , versión reducida o simetrizada del estimador e del parámetro θ , es como mínimo igual de preciso que e .*

El problema de la selección del “mejor” estimador, es decir, la elección del estimador $e(s, \mathbf{y})$ del parámetro $\theta(\mathbf{y})$ que minimice $E(e(s, \mathbf{y}) - \theta(\mathbf{y}))^2$ uniformemente en R^N unido a la condición de insesgadez origina la siguiente definición.

3.2.2 Estimadores UMVUE

Definición 3.11 *Un estimador $e(s, \theta(\mathbf{y}))$ basado en un diseño muestral d se dice que es uniformemente de mínima varianza para el parámetro $\theta(\mathbf{y})$ si y sólo si para todo estimador insesgado $e'(s, \theta(\mathbf{y}))$ de $\theta(\mathbf{y})$ se verifica:*

$$E(e(s, \mathbf{y})) = \theta(\mathbf{y}) \quad \forall \mathbf{y} \in R^N, \quad y$$

$$E(e(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \leq E(e'(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \quad \forall \mathbf{y} \in R^N.$$

3.3 Consistencia

Definición 3.12 *Un estimador se dice consistente cuando verifica que $e(U, \mathbf{y}) = \theta(\mathbf{y})$, $\forall \mathbf{y} \in R^N$.*

3.4 Suficiencia

El criterio de suficiencia va a tener un uso similar al que tiene en la inferencia estadística tradicional: reducir los datos observados a un conjunto más compacto sin perder información esencial en el problema de inferencia en cuestión. Así, la búsqueda de estimadores podrá restringirse a la clase de estimadores que sean función de un estadístico suficiente.

Definición 3.13 *En un diseño ordenado d^* un estadístico $z = e(s^*, \mathbf{y}) = e(\mathcal{D}^*)$ se dice que es suficiente para el vector paramétrico $\mathbf{y} \in R^N$ si y sólo si la distribución de \mathcal{D}^* condicionada a que $Z = z$ no depende de \mathbf{y} .*

Ejemplo 3.14 *La función de reducción de los datos, r , que permite disminuir la información de una secuencia de unidades de una muestra s^* (con unidades posiblemente ordenadas y con multiplicidad) a otra muestra s con unidades distintas, $r(s^*) = s$, con $r : S_d^* \rightarrow S_d$, también permite pasar del diseño ordenado d^* al diseño no ordenado d . El estadístico $\mathcal{D} = r(\mathcal{D}^*)$ definido como $r(\delta^*)$ cuando \mathcal{D}^* toma el valor δ^* , es un estadístico suficiente para \mathbf{y} (y minimal suficiente).*

3.5 Completitud

Definición 3.15 *En un diseño ordenado d^* un estadístico $z = e(s^*, \mathbf{y}) = e(\mathcal{D}^*)$ se dice que es completo si y sólo para cada función $g(z)$ con $E(g(z)) = 0 \forall \mathbf{y} \in R^N$, se tiene que $P(g(z) = 0) = 1 \forall \mathbf{y} \in R^N$.*

Así, el estadístico función de reducción de los datos no es completo, pues de serlo sería UMVUE y como veremos más adelante esto no ocurre.

4 Estrategias

Como la distribución en el muestreo de un estimador $e(s, \mathbf{y})$ de un parámetro $\theta(\mathbf{y})$ viene determinada por la función de probabilidad definida sobre el soporte del diseño, P_d , si la muestra se extrae con un diseño distinto, su distribución también cambiará, y las buenas

propiedades con las que se seleccionó el estimador en un diseño d (insesgadez y máxima precisión) pueden no cumplirse en otro diseño distinto. Surge así un concepto nuevo que se define a continuación.

Definición 4.1 *Una estrategia muestral para un parámetro $\theta(\mathbf{y})$ de una población finita \mathcal{U} es un par, (d, e) formado por un diseño muestral $d = (S_d, P_d)$ y un estimador $e(s, \mathbf{y})$.*

La comparación de estimadores antes expuesta puede generalizarse para estrategias, como haremos a continuación.

4.1 Comparación de estrategias

Definición 4.2 *Dadas dos estrategias $(d_1, e_1(s, \theta(\mathbf{y})))$ y $(d_2, e_2(s, \theta(\mathbf{y})))$ para el parámetro $\theta(\mathbf{y})$, se dice que (d_1, e_1) es “mejor” que (d_2, e_2) si y sólo si*

$$E_{d_1}(e_1(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \leq E_{d_2}(e_2(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \quad \forall \mathbf{y} \in R^N.$$

El problema de determinar una estrategia mejor que otras no tiene solución, como veremos más adelante, lo que da sentido a la definición equivalente a la definición que se hizo en el caso de los estimadores, la admisibilidad.

Las comparaciones entre estrategias sólo podrán efectuarse para un $\mathbf{y} \in R^N$ fijo, lo que en la práctica da sentido a que una vez esté fijada la población finita se intenten buscar combinaciones de estimadores y diseños muestrales con el fin de ser comparados y seleccionar aquél de ellos que produzca estimaciones mejores de los parámetros de interés.

5 Resultados notables

5.1 Teorema de Basu

Teorema 5.1 *Teorema de Basu. Para un diseño muestral que no sea un censo y para un parámetro $\theta(\mathbf{y})$ que dependa de los valores de $\mathbf{y} \in R^N$, no existe un estimador UMVUE.*

El teorema de Basu (1971) proporciona uno de los resultados más importantes y característicos de la inferencia en muestreo de poblaciones finitas: la no existencia de estimadores UMVUE entre la clase de estimadores insesgados de un parámetro que depende de los valores de y .

Así, la búsqueda del “mejor” estimador para por ejemplo el parámetro Y , no tiene solución para la clase de estimadores insesgados, pero se pueden obtener algunos resultados si esta búsqueda se restringe a una clase de estimadores más pequeña.

A continuación vamos a dar algunos resultados para la clase C_L de estimadores lineales homogéneos e insesgados del total Y . En dicha clase se encuentra el estimador \hat{Y}_{HT} , que como veremos ocupa un papel destacado.

Teorema 5.2 *Para un diseño unicluster, el estimador de Horvitz–Thompson es el único mejor estimador de Y en la clase C_L .*

Demostración.-

Sea $l(s, \mathbf{y})$ un estimador lineal homogéneo e insesgado de Y :

$$l(s, \mathbf{y}) = \sum_{i \in s} l_{s_i} y_i, \quad s \in S.$$

Consideramos su versión simetrizada $l^*(s, \mathbf{y})$ que es también lineal y homogéneo:

$$l^*(s, \mathbf{y}) = \sum_{i \in s} l_{s_i}^* y_i, \quad \text{con} \quad l_{s_i}^* = l_{s_i^*}^* \quad \forall i \in s, s^*, s \approx s^*.$$

Puesto que el diseño es unicluster, dadas dos muestras s' y s'' , o $s' \cap s'' = \emptyset$ o bien $s' \approx s''$. Como el estimador es insesgado:

$$\begin{aligned} E(l^*(s, \mathbf{y})) &= \sum_s \left(\sum_{i \in s} l_{s_i}^* y_i \right) p(s) = \sum_s y_i \left(\sum_{i \in s} l_{s_i}^* p(s) \right) = \\ &= \sum_s y_i l_{s_i}^* \left(\sum_{i \in s} p(s) \right) = \sum_s y_i l_{s_i}^* \pi_i = Y \end{aligned}$$

uniformemente en $\mathbf{y} \in R^N$, por lo que $l_{s_i}^* = \pi_i^{-1}$, y así $l^*(s, y) = \sum_{i \in s} \frac{y_i}{\pi_i}$, $\forall s \in S_d$.

Entonces la versión simetrizada de cada estimador de la clase C_L de Y es única, $l^*(s, \mathbf{y}) = \hat{Y}_{HT}$, que es por tanto el mejor estimador de Y en la clase C_L .

Este resultado de optimalidad del estimador de Horvitz–Thompson en la clase C_L se obtiene si el diseño muestral es unicluster. Teniendo en cuenta que los diseños muestrales más usados no son de este tipo, nos planteamos el estudio del mismo problema para un diseño no unicluster, cuyo resultado es bastante distinto.

Teorema 5.3 *Para un diseño no unicluster, no existe ningún mejor estimador de Y en C_L .*

La demostración original de este teorema fue dada por Godambe (1955) y Hanurav (1966). Una demostración posterior más ilustrativa puede verse en Hedayat y Sinha (1991) (pág 36–37).

Así pues dado un diseño arbitrario, no se tiene garantizada la existencia del mejor estimador, ni siquiera en la clase C_L .

Ejemplo 5.4 *Sea una población $\mathcal{U} = \{1, 2, 3, \dots, 10\}$ y sobre ella los diseños muestrales d_1 con soporte*

$$S_d = \{(1, 2), (3, 4, 7), (5, 6, 9), (8, 10)\}$$

y probabilidad

$$P_{d1}(1, 2) = 0.2, P_{d1}(3, 4, 7) = 0.3, P_{d1}(5, 6, 9) = 0.4, P_{d1}(8, 10) = 0.1$$

y d_2 con soporte

$$S_d = \{(2, 4, 7), (9, 6, 3, 1), (10, 5, 8), (5, 8, 10), (1, 3, 6, 9)\}$$

con

$$P_{d2}(2, 4, 7) = 0.2, P_{d2}(9, 6, 3, 1) = 0.5, P_{d2}(1, 3, 6, 9) = 0.1$$

$$P_{d2}(10, 5, 8) = 0.1 = P_{d2}(5, 8, 10)$$

La forma reducida de d_2 es d_2^ , y entonces mejor estimador de la media es:*

bajo d_1 :

$$\bar{y}_{HT}(1, 2) = \frac{5}{10}(y_1 + y_2), \bar{y}_{HT}(3, 4, 7) = \frac{10}{30}(y_3 + y_4 + y_7), \bar{y}_{HT}(5, 6, 9) = \frac{5}{20}(y_5 + y_6 + y_9) \text{ y } \bar{y}_{HT}(8, 10) = \frac{10}{10}(y_8 + y_{10}),$$

bajo d_2^* :

$$\bar{y}_{HT}(2, 4, 7) = \frac{5}{10}(y_2 + y_4 + y_7), \bar{y}_{HT}(5, 8, 10) = \frac{5}{10}(y_5 + y_8 + y_{10}) \text{ y } \bar{y}_{HT}(1, 3, 6, 9) = \frac{50}{30}(y_1 + y_3 + y_6 + y_9)$$

puesto que los diseños son unicluster.

Recordando el concepto de admisibilidad dado con anterioridad, un estimador e_2 se dice inadmisble dentro de la clase de estimadores insesgados de θ , si existe otro estimador “mejor” que él. Por el contrario, un estimador se dice admisible si no puede ser mejorado uniformemente por otro estimador, es decir, $e(s)$ es admisible si y sólo si $\forall e'$ insesgado de $\theta(\mathbf{y})$,

$$E(e(s, \mathbf{y}) - \theta(\mathbf{y}))^2 = E(e'(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \quad \forall \mathbf{y} \in R^N$$

o bien

$$\exists W \subset R^N \mid E(e(s, \mathbf{y}) - \theta(\mathbf{y}))^2 < E(e'(s, \mathbf{y}) - \theta(\mathbf{y}))^2 \quad \forall \mathbf{y} \in W.$$

Si un estimador es admisible no está injustificado el utilizarlo para inferir $\theta(\mathbf{y})$. Por tanto la admisibilidad se puede ver como un mínimo requerimiento que se espera para un buen estimador.

Nos planteamos así en un diseño no unicluster, la posible existencia de un estimador admisible, y, en tal caso, su unicidad. Los siguientes resultados debidos a Godambe (1960) y Joshi (1965) se refieren a la admisibilidad del estimador de Horvitz–Thompson.

Teorema 5.5 *El estimador \hat{Y}_{HT} es admisible en la clase C_L de estimadores lineales homogéneos e insesgados de Y .*

Demostración.

Consideremos un estimador $l(s, y) \in C_L$, $l(s, y) = \sum_{j \in s} l_{sj} y_j$, verificando la condición de insesgadez para Y :

$$\sum_{s \ni i} l_{is} P_d(s) = 1, \forall i \in s$$

y sean los vectores: $y_{(i)} = (0, 0, \dots, y_i, 0, \dots, 0)^t$, ($1 \leq i \leq N$) con $y_i \neq 0$ entonces:

$$V(l(s, y)) \mid_{y=y_{(i)}} = y_i^2 \left(\sum_{s \ni i} l_{is}^2 P_d(s) - 1 \right)$$

Si consideramos el estimador de H-T,

$$V(\hat{Y}_{HT}) \mid_{y=y_{(i)}} = y_i^2 \left(\frac{1}{\pi_i} - 1 \right).$$

La desigualdad de Cauchy-Schwartz permite deducir que:

$$\sum_{s \ni i} l_{is}^2 P_d(s) \cdot \sum_{s \ni i} P_d(s) \geq \left(\sum_{s \ni i} l_{is} P_d(s) \right)^2$$

y teniendo en cuenta la condición de insesgadez, y la definición de π_i se tiene que:

$$\sum_{s \ni i} l_{is}^2 P_d(s) \geq \pi_i^{-1}$$

De esta desigualdad se deduce que:

$$V(l(s, y)) \mid_{y=y_{(i)}} \geq V(\hat{Y}_{HT}) \mid_{y=y_{(i)}}$$

alcanzándose la igualdad sí y sólo si $l_{si} = \pi_i^{-1}$, $\forall i \in s$, tal que $P_d(s) > 0$.

Finalmente si definimos el conjunto $W = \cup_{i=1}^N \{y_{(i)} \mid y_i = 0\}$ se obtiene que

$$V(l(s, y)) \geq V(\hat{Y}_{HT}), \forall i \in W$$

con desigualdad estricta para al menos un punto de W , a no ser que coincidieran los coeficientes con π_i^{-1} en cuyo caso coincidiría con el estimador de H-T. En consecuencia se tiene la admisibilidad del estimador de H-T en la clase de estimadores lineales y homogéneos insesgados para el total.

Teorema 5.6 *El estimador \hat{Y}_{HT} es admisible en la clase de los estimadores insesgados de Y .*

Las demostraciones de este teorema es más compleja y se pueden encontrar en el texto de Hedayat y Sihna (1991).

Corolario 5.7 *En $MAS(N, n)$ el estimador usual $\hat{Y} = N\bar{y}$ es admisible en la clase de estimadores insesgados de Y .*

5.2 El teorema de Rao–Blackwell

Vamos a ver en este apartado que el estadístico suficiente $\mathcal{D} = r(\mathcal{D}^*)$ se puede utilizar para la estimación bajo diseños ordenados. El principal resultado será la adaptación del teorema de Rao–Blackwell al muestreo en poblaciones finitas.

Sea δ^* un dato ordenado y sea $t = t(\mathcal{D}^*)$ un estimador no necesariamente insesgado del parámetro $\theta(\mathbf{y})$. Para cada valor $\delta = r(\delta^*)$ de la variable $\mathcal{D} = r(\mathcal{D}^*)$ se define

$$t^*(\delta) = E(t(\mathcal{D}^*) \mid \mathcal{D} = \delta).$$

Puesto que \mathcal{D} es suficiente para \mathbf{y} , la función t^* no depende de parámetros desconocidos, y por tanto $t^* = t^*(\mathcal{D})$ es un estadístico que depende de \mathcal{D}^* sólo a través de $\mathcal{D} = r(\mathcal{D}^*)$.

Teorema 5.8 *Sea $t = t(\mathcal{D}^*)$ un estimador de $\theta(\mathbf{y})$ y sea $t^* = t^*(\mathcal{D})$ definido anteriormente. Entonces:*

- i) $E(t) = E(t^*)$,
- ii) $ECM(t) = ECM(t^*) + E((t - t^*)^2)$, y
- iii) $ECM(t^*) \leq ECM(t)$.

El estimador t^* se ha construido por el proceso de Rao–Blackwelización del estimador original t . Puesto que t^* depende del dato sólo a través de δ , t^* es independiente del orden y multiplicidad de las unidades de la muestra. Este teorema garantiza cómo obtener un estimador mejor que uno dado que dependa del orden y multiplicidad de las unidades de la muestra, que no dependa de estos factores.

Nos restringiremos en adelante a estimadores que dependan de \mathcal{D}^* sólo a través de $\mathcal{D} = r(\mathcal{D}^*)$, y así, la búsqueda de estimadores con buenas propiedades se hará entre aquellos estimadores que no dependan del orden o la multiplicidad de las unidades.

Ejemplo 5.9 Un estimador de la media de una variable y en un diseño con un soporte cuyas muestras, s^* , contengan elementos con multiplicidad, $\bar{y}_{s^*} = \sum_{i \in s^*} \frac{y_i}{n(s^*)}$ puede mejorarse con $t^* = E(\bar{y}_{s^*} | D = \delta) = \sum_{i \in s} \frac{y_i}{\nu(s)}$, siendo $\nu(s)$ el tamaño muestral efectivo.

5.3 El criterio de máxima verosimilitud

En muestreo de poblaciones finitas es sencillo de identificar la función de verosimilitud, si bien no será un criterio operativo en los problemas de inferencia ya que no se obtendrá un único estimador.

Definición 5.10 Para un dato δ , se define su función de verosimilitud, $L_\delta(\mathbf{y})$, como una función del parámetro \mathbf{y} tal que $\forall \mathbf{y} \in R^N$ da la probabilidad de obtener δ si es \mathbf{y} el verdadero valor del parámetro.

Teorema 5.11 Para cada diseño $d = (S_d, P_d)$, la función de verosimilitud viene dada por

$$L_\delta(\mathbf{y}) = \begin{cases} P_d(s) & \forall \mathbf{y} \in \Omega_\delta \\ 0 & \text{en otro caso} \end{cases}$$

donde $\Omega_\delta = \{\mathbf{y} \in R^N \mid \mathbf{y} \text{ es consistente con } \delta\}$.

Al ser la función de verosimilitud, $L_\delta(\mathbf{y})$ constante $\forall \mathbf{y} \in \Omega_\delta \subset R^N$ no tiene un único máximo y en consecuencia el principio de máxima verosimilitud no proporciona un único estimador del parámetro \mathbf{y} .

Ejemplo 5.12 Sea $N = 4$, la muestra obtenida $s = (1, 2)$ con $P(s) = 0.2$ y el dato $\delta = ((1, 20), (2, 0))$. Entonces, $\Omega_\delta = \{(20, 0, y_3, y_4) \mid -\infty < y_3, y_4 < +\infty\} \subset R^4$ y todos los posibles valores del vector $(20, 0, y_3, y_4)$ tienen la misma verosimilitud $L_\delta((20, 0, y_3, y_4)) = 0.2$ y no existe información suficiente para producir un único estimador máximo verosímil de, por ejemplo, $\bar{Y} = (20 + 0 + y_3 + y_4)/4$.

El principio de máxima verosimilitud afirma que si dos procesos de obtención de datos proporcionan funciones de verosimilitud que difieren sólo en una constante (constante respecto al parámetro desconocido), entonces los dos procedimientos proporcionan la misma inferencia. Así, en el muestreo en poblaciones finitas, dados dos

diseños d_1 y d_2 para los que exista una muestra con $P_1(s) > 0$ y $P_2(s) > 0$, para cada dato será:

$$L_\delta^1(\mathbf{y}) = \begin{cases} P_1(s) & \forall \mathbf{y} \in \Omega_\delta \\ 0 & \text{en otro caso} \end{cases} \quad L_\delta^2(\mathbf{y}) = \begin{cases} P_2(s) & \forall \mathbf{y} \in \Omega_\delta \\ 0 & \text{en otro caso} \end{cases}$$

y entonces $L_\delta^1(\mathbf{y}) = \frac{P_1(s)}{P_2(s)} L_\delta^2(\mathbf{y})$, $\forall \mathbf{y} \in R^N$, es decir, las funciones de verosimilitud difieren en una constante, por lo que el dato δ proporciona la misma inferencia, según el principio de verosimilitud, para \mathbf{y} , independientemente del diseño del que proviene.

Este resultado es fundamental en el muestreo de poblaciones finitas, ya que, la inferencia a realizar a partir de los datos observados resulta independiente del diseño muestral. Este controvertido resultado no es aceptado por los expertos en muestreo de poblaciones finitas que consideran que las propiedades más importantes de los estimadores como la insesgadez o la precisión dependen del diseño muestral que se considere.