QÜESTIIÓ, vol. 23, 1, p. 3-39, 1999

UNA REVISIÓN DE DIFERENTES APORTACIONES AL DISEÑO EN POBLACIONES FINITAS

B. FONT
Universitat de València*

Este artículo es una revisión de los resultados más relevantes de diseño en poblaciones finitas. Los resultados presentados se clasifican en tres grupos: población fija, inferencia basada en modelos de superpoblación clásicos e inferencia basada en modelos de superpoblación Bayesianos, y se analizan de forma comparativa. Entre los resultados revisados señalemos las aportaciones sobre: procedimiento de muestreo, tamaño óptimo de la muestra y procedimientos de estratificación.

A comparative review of different contributions in survey sampling.

Palabras clave: Diseño óptimo, estrategias de muestreo, inferencia Bayesiana, modelos de superpoblación Bayesianos, modelos de superpoblación clásicos, población fija.

Clasificación AMS (MSC 2000): 62D05.

* Begoña Font Belaire. Dept. Economia Financera i Matemàtica. Edifici Departamental Oriental. Av. dels Tarongers s/n. 46071 València (Espanya).

– Recibido en agosto de 1997.
– Aceptado en enero de 1999.

1. INTRODUCCIÓN

La obtención de información a través de encuestas de la población se ha convertido en una práctica frecuente en la investigación de entidades públicas y privadas que demandan respuestas a preguntas tan básicas como: ¿cuál es el método más apropiado para obtener la muestra?, ¿cuál es el tamaño apropiado de la muestra?, ¿cómo debo analizar los datos obtenidos? Los que estudiamos poblaciones finitas sabemos que las respuestas a estas preguntas no son únicas y universales, y además dependen de la perspectiva de análisis aplicada. Al escribir este artículo se ha pensado que era importante realizar un esfuerzo en recoger y analizar las respuestas que da la literatura a estas preguntas de diseño desde las distintas perspectivas. De este modo, el lector no especializado obtendrá con este trabajo una visión conjunta sobre las distintas respuestas de la literatura a sus preguntas sobre diseño, y el lector especializado tendrá una referencia en la que se analizan conjuntamente los resultados sobre diseño desde las diversas perspectivas, con la particularidad añadida de permitir el estudio por separado de cada grupo de aportaciones por el carácter autocontenido de las secciones sobre aportaciones desde las perspectivas de población fija, modelos de superpoblación clásicos y modelos de superpoblación Bayesianos.

Resumiendo y concretando, este artículo realiza una revisión de la literatura clásica y reciente sobre diseño en poblaciones finitas desde las perspectivas de población fija, y modelos de superpoblación desde las perspectivas clásica y Bayesiana. En concreto, se revisan las aportaciones en relación a cómo seleccionar la muestra, elegir el tamaño óptimo de unidades y elementos a muestrear en un muestreo en dos etapas, técnicas para estratificación de una población, etc., señalando las hipótesis asumidas en cada caso desde las distintas perspectivas. Otros artículos de revisión sobre las diferentes aportaciones al diseño y estimación son: Zacks(1970), Rao(1979), y Bellhouse(1984).

Debemos indicar que en este artículo no se pretende realizar una revisión de las aportaciones realizadas en el terreno de obtener el estimador óptimo para un muestreo dado o para un grupo de procedimientos de muestreo (aunque se indiquen como referencia algunos resultados), y por lo tanto que la exposición no se centrará en la forma de los estimadores sino en las características relativas al diseño. El lector interesado en un análisis más detallado sobre estimación puede dirigirse, por ejemplo, a los siguientes artículos y libros:

– Sobre muestreo en población fija, a los manuales de Hansen, Hurwitz and Madow(1953), Kish(1965), Cochran(1977), Hedayat and Sinha(1991) y Särndal, Swenson and Wretman(1992) y a los artículos de Chaudhuri(1988), Mukhopadhyay and Tracy(1993) y Godambe(1992).

– Sobre muestreo y estimación basada en un modelo de superpoblación, a los manuales de Cassel, Särndal and Wretman(1976), Bolfarine and Zacks(1992) y a los artículos de Royall(1988,92a) desde las perspectivas clásica, y Ericson(1969a,88) y Font(1995b) desde la perspectiva Bayesiana.

Otras referencias de interés son el artículo de Särndal(1978) y las tesis de Murgui(1982) y Font(1995a) que recogen un resumen comparativo de las distintas aproximaciones al problema de estimación en poblaciones finitas.

El presente trabajo se organiza en 6 secciones incluyendo esta introducción. En la sección 2 se establecen los conceptos básicos y notaciones que se aplicarán en la descripción de los distintos resultados. En la sección 3 se presentan las aportaciones al diseño desde la perspectiva de población fija. En las secciones 4 y 5 las aportaciones al diseño que se apoyan en un modelo de superpoblación desde las metodologías clásica y Bayesiana respectivamente. Y en la sección 6 se presenta un resumen conclusión sobre la comparación realizada entre las distintas perspectivas y se referencian otras aportaciones al diseño no tratadas en las secciones 3, 4 y 5.

2. CONCEPTOS BÁSICOS Y NOTACIÓN

Consideremos una población finita de N (entero positivo conocido) elementos, $P = \{1, 2, \ldots, N\}$, unas cantidades y_i, $i = 1, 2, \ldots, N$ que representan el valor de una característica de interés asociada a cada uno de estos elementos, y supongamos que estamos interesados en inferir sobre la cantidad (desconocida) $\overline{y} = \frac{\sum_{i=1}^{N} y_i}{N}$. Para poder estimar $\overline{y}$ (denotaremos al estimador por $\widehat{\overline{y}}$), obtendremos una muestra $s = \{i_1, i_2, \ldots, i_n\}$ con probabilidad $p(s)$ mediante un mecanismo de muestreo no informativo de tamaño fijo n y mediremos para los índices muestreados el valor de la característica en estudio. En este artículo nos restringiremos a los resultados referidos a procedimientos de muestreo no informativos de tamaño fijo n. Dentro de este grupo indicaremos por srswr al muestreo aleatorio con reemplazamiento, ppswr al muestreo con reemplazamiento con $p_i = Pr(\text{seleccionar i})$, $i = 1, 2, \ldots, N$, srswor al muestreo aleatorio sin reemplazamiento, y ppswor al muestreo sin reemplazamiento con $\pi_i = Pr(i \in s)$, $i = 1, 2, \ldots, N$. A partir de esta notación básica introduciremos a continuación las notaciones particulares correspondientes a las perspectivas de población fija y modelos de superpoblación.

La perspectiva de población fija considera como única fuente de aleatoriedad la debida al mecanismo de muestreo. Si representamos por $E_p(.)$ y $V_p(.)$ a los operadores media y varianza respecto al procedimiento de muestreo p, diremos que nuestro estimador $\widehat{\overline{y}}$ es insesgado respecto al muestreo si:

$$E_p(\widehat{\overline{y}}) = \overline{y}, \tag{1}$$

y en este caso el error cuadrático medio cometido respecto al muestreo vendrá dado por:

$$E_p(\widehat{\overline{y}} - \overline{y})^2 = V_p(\widehat{\overline{y}}). \tag{2}$$

A partir de (1) y (2), Newman(1934) define como estrategia óptima (estrategia de uniforme mínima varianza) aquella estrategia $(\widehat{\overline{y}}, p)$ que minimice el error cuadrático medio respecto al muestreo p (ver (2)) entre todos los estimadores insesgados respecto a ese procedimiento de muestreo p (ver (1)). Los trabajos de Godambe(1955) y Godambe and Joshi(1965) demuestran la no existencia de estimadores óptimos según la definición anterior (más detalles en sección 3) y justifican la búsqueda de otros criterios de estrategia óptima (por ejemplo, la admisibilidad) o la introducción de modelos de superpoblación (ver otras razones en la sección 4) para delimitar en qué circunstancias relativas a la población un estimador era mejor que otro.

La perspectiva basada en modelos de superpoblación considera a los valores y_i como realizaciones de una variable aleatoria Y_i, $i = 1, 2, \dots, N$. Si asumimos esta segunda fuente de aleatoriedad y representamos por $E_m(.)$ y $V_m(.)$ a los operadores media y varianza respecto al modelo de superpoblación, diremos que el estimador $\widehat{\overline{y}}$ (manteniendo la notación inicial) es insesgado respecto al modelo si:

$$E_m(\widehat{\overline{y}} - \overline{Y}) = 0. \tag{3}$$

La literatura que considera un modelo de superpoblación propone dos criterios para seleccionar la estrategia óptima, según la importancia que el estadístico quiera dar a cada una de las dos fuentes de aleatoriedad (procedimiento de muestreo y modelo de superpoblación) que consisten en minimizar:

$$E_m E_p(\widehat{\overline{y}} - \overline{Y})^2, \tag{4}$$

o bien,

$$E_m(\widehat{\overline{y}} - \overline{Y})^2, \tag{5}$$

entre estimadores que satisfacen (1) ó (3).

A veces dispondremos también de una información complementaria sobre la variable en estudio y_i, en forma de una característica auxiliar x_i con $i = 1, 2, \dots, N$ conocida para todas las unidades de la población que se podrá aplicar en el mecanismo de muestreo y/o en la estimación para mejorar el proceso inferencial. Algunas notaciones adicionales de interés son:

$$\overline{y} = \frac{\sum_{i=1}^{N} y_i}{N}, \qquad \overline{y}_s = \frac{\sum_{i \in s} y_i}{n}, \qquad \overline{y}_u = \frac{\sum_{i \notin s} y_i}{N-n},$$

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}, \qquad \overline{x}_s = \frac{\sum_{i \in s} x_i}{n}, \qquad \overline{x}_u = \frac{\sum_{i \notin s} x_i}{N-n},$$

$$S_y^2 = \frac{\sum_{i=1}^{N} (y_i - \overline{y})^2}{N-1}, \quad S_x^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1}.$$

En algunas ocasiones los elementos de la población (de manera natural o bien tras un proceso de estratificación) aparecen agrupados en unidades más grandes (psu's). En este trabajo, denotaremos por K (entero positivo conocido) al número de unidades, M_i al número de elementos de la unidad i, $i = 1, 2, \ldots, K$ (enteros positivos conocidos) con $\sum_{i=1}^{K} M_i = N$, y_{ij} al valor de la característica en estudio asociada al elemento j de la unidad i, para $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, K$. Las inferencias que se analizarán en este trabajo se refieren a la media de la población por elemento y se apoyarán en los valores observados y_{ij} para $(i, j) \in s$, con s una muestra obtenida a través de un muestreo en dos etapas, en el que primero se realiza un muestreo no informativo de tamaño fijo k (denotaremos a la muestra obtenida por s_0) y en segundo lugar un muestreo no informativo de tamaño fijo m_i para $i \in s_0$ con $\sum_{i \in s_0} m_i = n$ (denotaremos a la muestra en cada unidad por s_i, $i \in s_0$) ambos procedimientos de muestreo independientes, por tanto $s = \{(i, j) : j \in s_i, i \in s_0\}$. En este trabajo indicaremos por: stsrswor al muestreo estratificado con $k = K$ con elección de los elementos mediante srswor, clsrswor al muestreo cluster con $k < K$ y $m_i = M_i$, $i = 1, 2, \ldots, K$ con elección de las unidades mediante srswor y 2ssrswor al muestreo en dos etapas con muestreo srswor de las unidades y de los elementos. Otras notaciones de interés son, para $i = 1, 2, \ldots, K$:

$$\overline{y}_i = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i}, \quad \overline{y}_{s_i} = \frac{\sum_{j \in s_i} y_{ij}}{m_i}, \qquad \overline{y}_{u_i} = \frac{\sum_{j \notin s_i} y_{ij}}{M_i - m_i},$$

$$\overline{x}_i = \frac{\sum_{j=1}^{M_i} x_{ij}}{M_i}, \quad S_{yi}^2 = \frac{\sum_{j=1}^{M_i} (y_{ij} - \overline{y}_i)^2}{M_i - 1}, \quad S_{xi}^2 = \frac{\sum_{j=1}^{M_i} (x_{ij} - \overline{x}_y)^2}{M_i - 1},$$

así como las siguientes definiciones de la varianza poblacional dentro de las unidades (S_{yw}^2), la varianza poblacional entre unidades (S_{yb}^2) y la correlación poblacional dentro de las unidades (R_y):

$$S_{yw}^2 = \frac{\sum_{i=1}^{K} (M_i - 1) S_{yi}^2}{N-K}, \quad S_{yb}^2 = \frac{\sum_{i=1}^{K} M_i (\overline{y}_i - \overline{y})^2}{K-1}, \quad R_y = \frac{2 \sum_{i=1}^{K} \sum_{j<j^*}^{M_i} (y_{ij} - \overline{y})(y_{ij^*} - \overline{y})}{\left(\frac{{}_{i=1}^{K} M_i^2}{N} - 1\right)(N-1) S_y^2},$$

esta última expresión (R_y), que puede interpretarse como un coeficiente de homogeneidad dentro de la unidad, ha sido aproximada por los diversos autores en términos de S_{yw}^2, S_{yb}^2 y S_y^2 (por ejemplo, en el caso $M_i = M$, $i = 1, 2, \ldots, K$, Hansen, Hurwitz and Madow(1953) usan la aproximación $\frac{\frac{M(K-1)}{N-1} S_{yb}^2 - S_y^2}{(M-1) S_y^2}$ y Cochran(1977) la aproximación $\frac{S_{yb}^2 - S_y^2}{(M-1) S_y^2}$). Indicaremos también con R_y a todas estas aproximaciones cuando formen parte de la aproximación de un resultado.

3. APORTACIONES AL DISEÑO DESDE LA PERSPECTIVA DE POBLACIÓN FIJA

Las aportaciones al diseño de los autores que han trabajado desde esta perspectiva son muy numerosas, en concreto en esta sección se analizarán algunos resultados comparando procedimientos de muestreo distintos, expresiones sobre los tamaños de muestreo óptimos y algunos procedimientos para estratificar la población con el objetivo de obtener mejores inferencias.

Antes de pasar a la relación de las aportaciones en población fija conviene realizar una reflexión previa sobre el concepto de estimador «óptimo», para valorar mejor las consecuencias de la elección entre las distintas alternativas de diseño en base a los errores cuadráticos asociados a ese estimador «óptimo». La no existencia de una estrategia uniforme de mínima varianza demostrada por Godambe(1955) cuando se consideraba la minimización dentro de la clase de los estimadores insesgados, lineales y homogéneos, y por Godambe and Joshi(1965) cuando se consideraba la clase de todos los estimadores insesgados (más detalles en los artículos de Chaudhuri(1988) y Mukhopadhyay and Tracy(1993)), conduce a establecer el concepto de estimador admisible para un muestreo dado. En esta línea de investigación, Joshi(1965,67) demostró: (i) la admisibilidad del estimador de Horwitz-Thompson ($\widehat{\overline{y}}_{HT} = \sum_{i\in s} \frac{y_i}{\pi_i N}$) como estimador de $\overline{y}$ para muestreos ppswor con probabilidades de inclusión π_i estrictamente positivas, (ii) la admisibilidad de $\overline{y}_s$ para un diseño p entre todos los estimadores de $\overline{y}$, y (iii) la admisibilidad de $\overline{y}_s$, $\widehat{\overline{y}}_r = \frac{\overline{x}}{\overline{x}_s}\overline{y}_s$ (estimador de razón), e $\widehat{\overline{y}}_b = \overline{y}_s + \widehat{\beta}(\overline{x} - \overline{x}_s)$ con $\widehat{\beta} = \frac{\sum_{i\in s}(x_i-\overline{x}_s)(y_i-\overline{y}_s)}{\sum_{i\in s}(x_i-\overline{x}_s)^2}$ (estimador de regresión) entre los estimadores medibles (con algunas restricciones adicionales) de $\overline{y}$ para funciones de pérdida muy generales para procedimientos de muestreo sin reemplazamiento. En otra línea de investigación, Newman(1934) y Royall(1968) demostraron que para un muestreo srswr, el estimador $\overline{y}_s$ es el único estimador que minimiza el error cuadrático respecto al muestreo entre los estimadores insesgados lineales homogéneos y entre todos los estimadores insesgados respectivamente.

3.1. Comparación entre procedimientos de muestreo

3.1.1. Comparación entre muestreos en una etapa

Veremos dos grupos de resultados, los debidos a Lanke(1975) que comparan los muestreos srswor y srswr y los de Des Raj(1954), Reddy and Rao(1977) y Rao(1993) que permiten comparar muestreos srswr y ppswr.

Lanke(1975) define que un procedimiento de muestreo p' es mejor o al menos tan bueno como otro procedimiento p si dado un estimador insesgado de $\overline{y}$ respecto a p, $\widehat{\overline{y}}_p$, existe un estimador insesgado de $\overline{y}$ respecto a p', $\widehat{\overline{y}}_{p'}$, tal que $V_{p'}(\widehat{\overline{y}}_{p'}) \leq V_p(\widehat{\overline{y}}_p)$, y obtiene una condición necesaria para que un muestreo p' sea mejor o al menos tan bueno como p para la estimación de $N\overline{y}$. De la aplicación de esta condición a los muestreos aleatorios se llega a que en la estimación de $\overline{y}$ un muestreo srswor es mejor o al menos tan bueno como un muestreo srswr si $n \geq 2$.

Des Raj(1954) demuestra que el muestreo ppxwr (muestreo ppswr con $p_i = \frac{x_i}{N\overline{x}}$, $i = 1, 2, \ldots, N$) puede producir peores estimaciones que el muestreo srswr cuando la relación entre y_i y x_i, $i = 1, 2, \ldots, N$ se aleja de una línea recta a través del origen.

Rao(1993), considera dos muestreos proporcionales con reemplazamiento, ppswr y pps'wr con probabilidades de elegir i, p_i y p_i', $i = 1, 2, \ldots, N$ respectivamente y considera el muestreo pps"wr con probabilidades de elegir i, $p_i'' = \alpha p_i + (1-\alpha)p_i'$ con $0 \leq \alpha \leq 1$, $i = 1, 2, \ldots, N$ obteniendo los siguientes resultados que representaremos por (1):

(i) Si $V_{\text{ppswr}}(\widehat{\overline{y}}_{HH}) < V_{\text{pps'wr}}(\widehat{\overline{y}}'_{HH})$, entonces $V_{\text{pps"wr}}(\widehat{\overline{y}}''_{HH}) < V_{\text{pps'wr}}(\widehat{\overline{y}}'_{HH})$.

(ii) Si $V_{\text{pps'wr}}(\widehat{\overline{y}}'_{HH}) < V_{\text{ppswr}}(\widehat{\overline{y}}_{HH})$, entonces $V_{\text{pps"wr}}(\widehat{\overline{y}}''_{HH}) < V_{\text{ppswr}}(\widehat{\overline{y}}_{HH})$.

(iii) $V_{\text{pps'wr}}(\widehat{\overline{y}}'_{HH}) - V_{\text{ppswr}}(\widehat{\overline{y}}_{HH}) = \frac{1}{nN^2}\sum_{i=1}^{N} \frac{y_i^2(p_i - p_i')}{p_i p_i'}$,

donde $\widehat{\overline{y}}_{HH} = \sum_{i \in s} \frac{y_i}{p_i N n}$ es el estimador de Hansen-Hurwitz asociado al muestreo ppswr (las primas representan los correspondientes estimadores Hansen-Hurwitz para los restantes muestreos).

De (1)(i) y (ii) se deduce que la estrategia $(\widehat{\overline{y}}''_{HH}, \text{pps"wr})$ es mejor que la peor entre las estrategias $(\widehat{\overline{y}}_{HH}, \text{ppswr})$ y $(\widehat{\overline{y}}'_{HH}, \text{pps'wr})$. De (1)(iii) tomando $p_i' = \frac{1}{N}$, $i = 1, 2, \ldots, N$ (muestreo srswr) obtenemos una expresión de las diferencias en eficiencia entre un muestreo srswr y un muestreo ppswr.

3.1.2. Comparación entre muestreos srswor, stsrswor y clsrswor

La comparación entre los muestreos srswor y stsrswor es analizada en la mayoría de manuales sobre técnicas de muestreo. Por ejemplo, Cochran(1977) y Särndal, Swensson and Wretman(1992) obtienen la siguiente expresión:

(2) $V_{\text{srswor}}(\overline{y}_s) - V_{\text{stsrswor,(13)}}(\overline{y}_s) = \frac{N-n}{nN(N-1)}[(K-1)S_{yb}^2 - \frac{1}{N}\sum_{i=1}^{K}(N-M_i)S_{yi}^2]$,

en la que el subíndice stsrswor (13) indica que se ha aplicado un muestreo aleatorio estratificado en el que el número de elementos muestreados en cada estrato es proporcional al tamaño del mismo (ver comentarios del apartado 3.2.2 y la expresión (13)).

A partir de (2) podemos observar que matemáticamente el muestreo stsrswor puede ser peor que el muestreo srswor. En particular, si suponemos que $S^2_{yi} = S^2_{psu}$, $i = 1, 2, \ldots, K$ tenemos que la expresión (2) se convierte en:

$$V_{\text{srswor}}(\overline{y}_s) - V_{\text{stsrswor,(13)}}(\overline{y}_s) = \frac{(N-n)(K-1)}{nN(N-1)}[S^2_{yb} - S^2_{psu}], \tag{3}$$

que será negativa (evidencia a favor del muestreo srswor) si la varianza poblacional entre estratos es más pequeña que la varianza dentro de los estratos. En la mayor parte de las situaciones prácticas la varianza entre estratos supera a la varianza dentro del estrato, (2) es positivo y se produce una ganancia en precisión al usar el muestreo stsrswor como alternativa al muestreo srswor.

Hansen, Hurwitz and Madow(1953) obtienen a partir de (2) la siguiente expresión aproximada de las ganancias derivadas del muestreo stsrswor frente al muestreo srswor:

$$V_{\text{srswor}}(\overline{y}_s) - V_{\text{stsrswor,(13)}}(\overline{y}_s) \simeq \frac{N-n}{N}\frac{S^2_y}{n}R_y. \tag{4}$$

A partir de (2) y (4) podemos extraer las siguientes observaciones:

– Cuanto más grande sea la homogeneidad dentro de los estratos, las ganancias por estratificación serán mayores.

– Cuanto más grande sea la heterogeneidad entre las medias de los estratos, mayores serán las ganancias relativas por estratificación.

– La ganancia relativa no será muy alta salvo que las variaciones entre estratos sean bastante mayores que las variaciones dentro de los estratos.

En relación a la comparación entre los muestreos srswor y clsrswor tenemos, apoyándonos en los resultados de Cochran(1977), cuando $M_i = M$, $i = 1, 2, \ldots, K$ que:

$$\begin{aligned} V_{\text{clsrswor}}(\overline{y}_s) - V_{\text{srswor}}(\overline{y}_s) &= \frac{N-n}{N}\frac{S^2_y}{n}\left\{\frac{N-1}{M(K-1)}[1 + (M-1)R_y] - 1\right\} \\ &\simeq \frac{N-n}{N}\frac{S^2_y}{n}(M-1)R_y. \end{aligned} \tag{5}$$

De (5) se deduce que si $R_y < 0$, el muestreo clsrswor es más eficiente que el srswor, y si $R_y > 0$ (la situación más habitual), el muestreo clsrswor es menos eficiente que el srswor. Hansen, Hurwitz and Madow(1953) proporcionan una buena discusión sobre los valores de R_y para diferentes ítems y diferentes tamaños de cluster y Särndal, Swensson and Wretman(1992) estudian los muestreos clppswor (muestreos de dos etapas con $m_i = M_i$, $i \in s_0$ con muestreo para las unidades del tipo ppswor) y aplicando un estimador Horwitz-Thompson llegan a las siguientes conclusiones prácticas:

- Si podemos seleccionar los clusters con un muestreo ppswor con $\pi_i = \frac{k\overline{y}_i}{\sum_{i=1}^{K} \overline{y}_i}$, $i = 1, 2, \ldots, K$, entonces este muestreo cluster será muy eficiente.
- Si seleccionamos los clusters con un muestreo ppswor con $\pi_i = \frac{kM_i}{N}$, $i = 1, 2, \ldots, K$, el muestreo será bastante eficiente si las variaciones entre las medias poblacionales de cada cluster son pequeñas.
- Un muestreo clsrswor es a menudo poco eficiente cuando los clusters son de distinto tamaño, es más, si la correlación entre M_i y $M_i\overline{y}_i$, $i = 1, 2, \ldots, K$ es positiva (que es el caso habitual), el muestreo clsrswor da peores resultados que en el caso de muestreo clsrswor con $M_i = M$, $i = 1, 2, \ldots, K$.

3.1.3. Comparación entre muestreos srswor y 2ssrswor

Hansen, Hurwitz and Madow(1953) obtienen la siguiente expresión para comparar los dos tipos de muestreo en el caso: $M_i = M$, $i = 1, 2, \ldots, K$ y $m_i = m$, $i \in s_0$

$$V_{\text{2ssrswor}}(\overline{y}_s) - V_{\text{srswor}}(\overline{y}_s) \simeq \frac{N-n}{N}\frac{S_y^2}{n}R_y(m-1). \tag{6}$$

A partir de (6) podemos concluir que un muestreo 2ssrswor será eficiente en la medida que la correlación entre los elementos dentro de cada unidad sea pequeña si es positiva, o en el medida que ésta sea negativa. En la práctica R_y es positiva y decrece al aumentar el tamaño de las unidades.

3.2. Tamaños de muestreo óptimos

La literatura clásica presenta tres criterios para determinar el tamaño de muestreo óptimo:

- Para un coste dado, determinar el tamaño que hace mínimo el error cuadrático medio en la estimación.
- Para una varianza respecto al muestreo dada del estimador de la media poblacional insesgado aplicado, determinar el tamaño que hace mínimo el coste de la estimación.
- El problema de decisión de minimizar una función de pérdida.

A continuación presentaremos los resultados obtenidos según estos criterios para muestreos srswor, stsrswor, clsrswor y 2ssrswor.

3.2.1. Tamaño de una muestra srswor

Konijn(1973) propone para una varianza V^2 dada la siguiente expresión para el tamaño n de una muestra srswor:

$$n = \frac{\frac{S_y^2}{V^2}}{\frac{N-1}{N} + \frac{S_y^2}{NV^2}}, \tag{7}$$

esta expresión cuando N es bastante grande se simplifica tomando $\frac{N-1}{N} \simeq 1$ a la expresión presentada por Kish(1965).

Y Cochran(1977) considerando el criterio de minimizar la pérdida cuadrática multiplicada por la constante positiva γ más una función lineal de costes $C(n) = c_0 + c_1 n$, obtiene la siguiente expresión para el tamaño óptimo n de una muestra srswor:

$$n = \frac{S_y^2 \sqrt{\gamma}}{\sqrt{c_1}}. \tag{8}$$

(Una forma más general de este resultado aparece en Yates(1960).) Debemos observar que las expresiones (7) y (8) proporcionan valores para n que son función de la varianza poblacional, e indicar que para poder fijar el tamaño óptimo la práctica habitual consiste en sustituir esta cantidad desconocida por un estimador.

3.2.2. Tamaño óptimo de una muestra stsrswor

Si consideramos la función de coste lineal: $C(m_1, \ldots, m_K) = c_0 + \sum_{i=1}^K c_i m_i$ se tiene (ver, por ejemplo, Särndal, Swensson and Wretman(1992)) que:

(i) Bajo la hipótesis de minimizar la varianza respecto al diseño para un muestreo stsrswor para un coste dado C, los tamaños óptimos m_i, $i = 1, 2, \ldots, K$ vienen dados por:

$$m_i = \frac{(C - c_0)\frac{M_i S_{yi}}{\sqrt{c_i}}}{\sum_{i=1}^K M_i S_{yi} \sqrt{c_i}} = n \frac{\frac{M_i S_{yi}}{\sqrt{c_i}}}{\sum_{i=1}^K \frac{M_i S_{yi}}{\sqrt{c_i}}}. \tag{9}$$

(ii) Bajo la hipótesis de minimizar el coste para una varianza dada V^2 para un muestreo stsrswor, los tamaños óptimos m_i, $i = 1, 2, \ldots, K$ vienen dados por:

$$m_i = \frac{\frac{M_i S_{yi}}{\sqrt{c_i}} \sum_{i=1}^K M_i S_{yi} \sqrt{c_i}}{(NV)^2 + \sum_{i=1}^K M_i S_{yi}^2}. \tag{10}$$

Observemos que a partir de (9) se obtiene tomando $c_i = c$, $i = 1, \ldots, K$, la asignación de Newman (el resultado sobre tamaño óptimo más popular), que consiste en la selección de muestras aleatorias estratificadas con tamaños de muestreo por estrato m_i, $i = 1, 2, \ldots, K$ dados por:

$$m_i = n \frac{M_i S_{yi}}{\sum_{i=1}^K M_i S_{yi}}. \tag{11}$$

Esta expresión fue obtenida por Neyman(1934) y previamente por Tschuprow(1923).

A partir de (9) y (10) se obtiene una regla práctica para diseñar un muestreo aleatorio estratificado que consiste en seleccionar, dado un estrato, una muestra más grande cuánto más grande sea su tamaño, más variable sea y menor sea el coste de muestrearlo, y las expresiones de los tamaños de los estratos y tamaño total de la muestra. Sin embargo, la obtención de estos tamaños presenta algunas dificultades prácticas, al desconocimiento de S_{yi}^2 (habitual en este tipo de expresiones y que discutiremos en el próximo párrafo) hay que añadir ahora otros problemas que se producen por la no introducción en los problemas de optimización resueltos en (9) y (10) de restricciones sobre los tamaños de la muestra en los estratos y que son: la posibilidad de que el tamaño óptimo de muestreo en cada estrato no sea entero, sea negativo o sea superior al tamaño del estrato. Las recomendaciones habituales son las siguientes: si el tamaño óptimo de muestreo supera al tamaño del estrato, muestrear estos estratos totalmente y distribuir las restantes observaciones entre los otros estratos, y si el tamaño óptimo es negativo no seleccionar ningún elemento de estos estratos, modificar la ecuación de coste y obtener los nuevos óptimos.

En la práctica, se suelen emplear las siguientes 5 aproximaciones de la fórmula (11) para evitar el problema que supone desconocer S_{yi}^2, $i = 1, 2, \ldots, K$:

(i) Asignación proporcional al total poblacional (que presenta problemas prácticos similares a (11) puesto que los totales de cada estrato son cantidades asimismo desconocidas)

$$m_i = n \frac{\sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^K \sum_{j=1}^{M_i} y_{ij}}. \tag{12}$$

(ii) Asignación proporcional al tamaño (que aplicábamos en (2) para establecer la comparación entre los muestreos srswor y stsrswor):

(13)
$$m_i = n\frac{M_i}{N}.$$

(iii) x-«óptima» asignación (con x una característica auxiliar que proporciona información acerca de y):

(14)
$$m_i = n\frac{M_i S_{xi}}{\sum_{i=1}^{K} M_i S_{xi}}.$$

(iv) Asignación proporcional a la raíz de las medias en cada estrato de una característica auxiliar:

(15)
$$m_i = n\frac{M_i\sqrt{\overline{x}_i}}{\sum_{i=1}^{K} M_i\sqrt{\overline{x}_i}}.$$

(v) Asignación proporcional al total de las x (con x una característica auxiliar que proporciona información acerca de y):

(16)
$$m_i = n\frac{\sum_{j=1}^{M_i} x_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{M_i} x_{ij}}.$$

La aplicación de las asignaciones (12) a (16) produce obviamente una pérdida en la precisión de las inferencias (en términos de la varianza) respecto a la aplicación de (11). Estas pérdidas de eficiencia dependerán en los casos (14) a (16) de la correlación entre la característica y en estudio y la característica auxiliar x (de valores conocidos para todos los elementos de la población) y en general de cuan buena sea la aproximación de S_{yi} por S_{xi}, $\sqrt{\overline{x}_i}$ o $\overline{x}_i$, $i = 1, 2, \ldots, K$ respectivamente. En el caso de la asignación proporcional (13) (la aproximación más empleada, porque no requiere conocer una característica auxiliar), Hendayat and Sinha(1991) obtienen la siguiente expresión:

(17)
$$V_{\text{stsrswor,(13)}}(\overline{y}_s) - V_{\text{stsrswor,(11)}}\left(\frac{{}^{K}_{i=1} M_i\overline{y}_{s_i}}{N}\right) = \frac{1}{nN}\sum_{i=1}^{K} M_i\left(S_{yi} - \frac{{}^{K}_{i=1} M_i S_{yi}}{N}\right)^2,$$

de la que se deduce que las diferencias de eficiencia entre las asignaciones de Newman y proporcionales pueden ser considerables, en especial si las varianzas dentro de cada estrato son muy diferentes entre sí.

3.2.3. Tamaño óptimo de una muestra clsrswor

La literatura no presenta muchos resultados en este campo y propone aplicar bajo la hipótesis de un coste cero para el muestreo dentro de las unidades los resultados que se presentaron en el apartado dedicado al tamaño para un muestreo srswor. Cochran(1977) para muestreos clsrswor con todos los clusters del mismo tamaño M propone una solución al problema de minimizar la varianza para un coste fijo con función de costes $C(k) = c_1 M k + c_2 \sqrt{k}$.

3.2.4. Tamaño óptimo de una muestra 2ssrswor

Para presentar los resultados distinguiremos entre dos situaciones:

– Todas las unidades tienen el mismo número de elementos: $M_i = M, i = 1, 2, \ldots, K$ y $m_i = m$ para $i \in s_0$.

– No todas las unidades tienen el mismo número de elementos.

Suponiendo que todas las unidades son del mismo tamaño, Cochran(1977) estudia el problema de minimizar la varianza para un coste dado, con función de costes lineal de la forma: $C(m, k) = c_1 k + c_2 k m$ y propone como método para obtener los tamaños m y k óptimos el siguiente algoritmo sugerido por Eisenhart (ver Cameron(1951)).

Algoritmo (18). Dados $a = c_1 S_{yw}^2$ y $b = \frac{c_2}{M}(S_{yb}^2 - S_{yw}^2)$:

(i) Si $b > 0$ y $\sqrt{\frac{a}{b}} \leq M$, tomar el entero l que cumpla $l \leq \sqrt{\frac{a}{b}} \leq l + 1$ y entonces:

$$\begin{cases} \text{si } \frac{a}{b} \leq l(l+1) & \text{tomar } m = l \\ \text{en otro caso} & \text{tomar } m = l + 1 \end{cases}$$

(ii) Si $b \leq 0$ ó $\sqrt{\frac{a}{b}} > M$, tomar $m = M$.

(iii) Determinado m óptimo, se calcula k fijando el coste o la varianza.

Nuevamente, la determinación del tamaño óptimo depende de cantidades desconocidas (varianza entre los elementos dentro de las unidades y varianza entre las medias de las unidades) aunque en este caso los tamaños óptimos son enteros y positivos. Hedayat and Sinha(1991) se plantean nuevamente el problema de minimizar la varianza para un coste menor o igual a una cantidad fijada (la función de costes estudiada es prácticamente la misma, sólo añaden un coste fijo) y sujeta a la restricción $2 \leq m \leq M$ y $2 \leq k \leq K$ presentando un algoritmo alternativo para la obtención de los tamaños muestrales óptimos m y k.

En el caso más general, en el que el tamaño de todas las unidades no es el mismo, Cochran(1977) asumiendo la función de costes $C(\overline{m}, k) = c_1 k + c_2 k\overline{m}$ donde $\overline{m} = \frac{n}{k}$ y considerando una fracción de submuestreo constante para todas las unidades muestreadas ($\frac{m_i}{M_i} = \overline{m}\frac{K}{N}$, $i \in s_0$) propone aplicar el algoritmo (18) reemplazando m por $\overline{m}$ y tomando $a = c_1 \frac{\sum_{i=1}^{K} M_i S_{yi}^2}{N}$ y $b = c_2(\frac{K^2 \sum_{i=1}^{K} M_i^2 (\overline{y}_i - \overline{y})^2}{N^2(K-1)} - \frac{K \sum_{i=1}^{K} M_i S_{yi}^2}{N^2})$. La solución obtenida $\overline{m}$ sirve para obtener los tamaños óptimos m_i, $i \in s_0$ aplicando que la fracción de submuestreo es constante para todas las unidades muestreadas.

Otra alternativa propuesta por Hansen, Hurwitz and Madow(1953) para la misma función de coste consiste en tomar, para $i \in s_0$:

$$m_i = M_i \frac{K}{N} \sqrt{\frac{c_1}{c_2} \frac{1 - R_y}{R_y}}, \tag{19}$$

y calcular k fijando el costo o la varianza. (La expresión (19) es asimismo aplicable cuando todas las unidades tiene el mismo tamaño.)

Otras referencias adicionales de interés son: Cochran(1977), Konijn(1973) y Hansen, Hurwitz and Madow(1953) con funciones de coste alternativas y/o análisis de otros procedimientos de muestreo en dos etapas.

3.3. Técnicas de estratificación de una población

Desde un punto de vista de población fija, el muestreo stsrswor permite obtener mejores inferencias que los muestreos srswor y clsrswor, y sobre todo cuando además tomamos la muestra de acuerdo con un esquema de asignación óptima. Esta propiedad motiva la búsqueda de procedimientos para dividir los elementos de la población en unidades (cuando la población no está dividida de forma natural) o incluso estratificar a posteriori la población de modo que se puedan aprovechar las ganancias en eficiencia del muestreo stsrswor frente al srswor o mejorar los resultados de un muestreo clsrswor. Trataremos en este apartado de dos cuestiones: cómo construir los estratos y el número de estratos.

La idea más sencilla para construir K estratos con el objetivo de mejorar nuestras estimaciones consiste en establecer una partición en estratos, de modo que las diferencias $M_i S_{iy}^2$ entre los estratos que construyamos no sean muy grandes, de esta manera cuando realicemos el muestreo de los mismos con asignación proporcional al tamaño (por ejemplo), las pérdidas de eficiencia respecto a la asignación de Neyman (ver expresión (17)) serán menores y, en consecuencia, nuestra estimación será mejor. Som(1973) considera tres procedimientos para reducir las diferencias $M_i S_{iy}^2$ entre estratos:

– La regla de Dalenius and Gurney(1951) de construir estratos con valores de $M_i S_{iy}^2$ aproximadamente iguales.

– La regla de Ekman(1959) de construir estratos con valores de $M_i\text{rg}_i$ aproximadamente iguales, con rg_i el rango de la característica en estudio en el estrato i, $i = 1, 2, \ldots, K$.

– Y la regla de Mahalanobis, Hansen, Hurwitz y Madow de construir estratos con valores de $M_i\overline{y}_i$ aproximadamente iguales.

Otra posibilidad consiste en hacer una estratificación de la población de modo que la varianza del estimador para la asignación de Pearson (ver (11)) sea directamente pequeña. Si representamos por $f(y)$ a una función de frecuencia sobre la característica en estudio y suponemos ordenada la población de acuerdo con esta característica, una manera de construir los estratos atendiendo a esta idea consistiría en determinar los límites intermedios $[y^0_{i-1}, y^0_i[$ de modo que se minimizara $V_{\text{stsrswor,(11)}}(\frac{\sum_{i=1}^{M_i} M_i\overline{y}_{s_i}}{N}) \simeq \sum_{i=1}^{K} \omega_i^2 \frac{\sigma_i^2}{m_i}$ con $\omega_i = \int_{y^0_{i-1}}^{y^0_i} f(y)dy$ y $\sigma_i^2 = \frac{1}{\omega_i} \int_{y^0_{i-1}}^{y^0_i} y^2 f(y)dy - [\frac{1}{\omega_i} \int_{y^0_{i-1}}^{y^0_i} y f(y)dy]^2$. Dalenius and Hodges(1959) plantean el problema en estos términos y proponen calcular la cumulativa de $\sqrt{f(y)}$ y tomar los límites intermedios de forma que los intervalos sean aproximadamente iguales en la escala cum$\sqrt{f(y)}$. (La idea de emplear la cumulativa $\sqrt{f(y)}$ y estos límites se apoya en la hipótesis de que si los estratos son numerosos y de rango pequeño, $f(y)$ dentro del estrato es aproximadamente constante.) Otros trabajos en esta línea (aunque en estos casos se propone muestrear un único elemento por estrato) consisten en la aplicación de los algoritmos de Kossack and Shiledar Baxi(1971) y sus modificaciones en Shiledar Baxi(1982,95).

Al enunciar los criterios de estratificación anteriores nos hemos apoyado en los valores de la característica en estudio, esta hipótesis no es realista ya que no conocemos la totalidad del vector poblacional. En la práctica estos métodos se aplican sobre una característica auxiliar x_i, $i = 1, 2, \ldots, N$ y, obviamente, la eficiencia en la construcción depende de la correlación entre la característica en estudio y la auxiliar.

Los procedimientos anteriores se apoyan en construir un número dado, K de estratos. En principio, un aumento de K supondrá (basándonos en un mecanismo de construcción bueno) estratos internamente más homogéneos en sí y más hetereogéneos entre sí y por tanto apoyándonos en (3) y (17) mayores ganancias de precisión. En esta dirección, Cochran(1977) apoyándose en una construcción de estratos basada en la regla de Dalenius and Hodges(1959) y suponiendo una distribución $f(y)$ dentro del estrato aproximadamente constante, obtiene la siguiente relación:

$$V_{\text{stsrswor}}(\frac{\sum_{i=1}^{M_i} M_i\overline{y}_{s_i}}{N}) \simeq \frac{V_{\text{srswor}}(\overline{y}_s)}{K^2}, \tag{20}$$

de modo que la varianza estratificada es inversamente proporcional al cuadrado del número de estratos. Por otra parte, hay razones a favor de un K no muy alto, así cuando para la construcción de los estratos se emplea una característica auxiliar (que es lo habitual en la práctica) existe un tamaño K' a partir del cual no es posible reducir la varianza por estratificación (este tamaño tope se debe a la varianza de la característica en estudio no explicable por la característica auxiliar). Otro motivo es el coste, mayor número de estratos supone mayores costes por trabajos extra en planificación y cálculo, y estos costes adicionales pueden no justificar las reducciones consiguientes (si se producen) de la varianza.

4. APORTACIONES AL DISEÑO DESDE LA PERSPECTIVA DE MODELOS DE SUPERPOBLACIÓN CLÁSICOS

Las aportaciones desde la perspectiva de modelos de superpoblación se apoyan en considerar que la característica en estudio y_i es realización de una variable aleatoria Y_i, $i = 1, 2, \ldots, N$ que se distribuye según un determinado modelo probabilístico. Desde el punto de vista del diseño podemos señalar dos grupos de motivaciones para introducir esta nueva fuente de aleatoriedad. Desde la perspectiva de población fija, la introducción del modelo de superpoblación permite la obtención de estrategias óptimas de diseño definidas, como aquellas que se obtienen minimizando el valor esperado respecto al modelo del error cuadrático medio respecto al diseño (minimizar (sección 2,4)), dentro de la clase de estimadores insesgados respecto al mecanismo de muestreo (ver expresión (sección 2,1)). Y permite valorar en presencia de información auxiliar en forma de una covariable x_i, $i = 1, 2, \ldots, N$ conocida, la eficiencia de los estimadores de razón o regresión frente al estimador media, y la eficiencia de las aproximaciones a los tamaños óptimos de la muestra y de construcción de estratos basados en la información proporcionada por esta covariable. Por otra parte, desde una perspectiva predictiva, otro grupo importante de autores defienden que el conocimiento del estadístico sobre las características de la población en estudio le permite argumentar que la población finita es la realización hoy de un determinado modelo de superpoblación o la realización de un modelo que describe un mecanismo o proceso en el mundo real; afirman que el modelo asumido constituye la fuente de aleatoriedad más importante ya que describe la población y definen como estimador óptimo (estimador óptimo desde la perspectiva predictiva) a aquel que minimiza el error cuadrático respecto al modelo (minimiza (sección 2,5)) dentro de la clase de estimadores insesgados respecto al modelo (ver sección 2,3). (Notemos que, para un muestreo no informativo es equivalente minimizar (sección 2,4) o (sección 2,5) para una clase dada de estimadores.)

Las diferencias conceptuales entre los defensores de cada una de estas perspectivas son importantes. De una manera resumida, los argumentos esgrimidos por la perspectiva población fija-modelo de superpoblación son que al realizar el análisis, la población está fijada, no es resultado de un proceso aleatorio y en consecuencia la fuente de aleatoriedad está en el mecanismo de selección de los elementos de las muestras sobre los que vamos a basar nuestra estimación, y que la estimación predictiva está sujeta a mayores sesgos en la estimación y mayores errores por especificación incorrecta del modelo. Los argumentos respuesta de los autores predictivos son que el modelo de superpoblación tiene las mismas características de objetividad y de interpretación en términos de frecuencias que los modelos considerados de forma habitual en los trabajos estadísticos en poblaciones infinitas, que la circunstancia de que las observaciones de la población finita ya estén realizadas aunque siguen siendo desconocidas para nosotros, no convierte en menos apropiada la aplicación del modelo de superpoblación que si hubieramos considerado el mismo modelo pero antes de que la población finita fuera realizada, y la realización de estudios sobre estimadores robustos frente a errores en la especificación del modelo.

En esta sección distinguiremos entre las aportaciones respecto al diseño realizadas desde las dos perspectivas anteriores: aportaciones desde la metodología población fija-superpoblación y desde la perspectiva predictiva, y nos apoyaremos en los siguientes dos modelos:

- **Modelo Clásico Base** (C$_1(\beta_0, \beta_1; v(x_i), \rho)$):

  $$Y_i = \beta_0 + \beta_1 x_i + E_i, \quad i = 1, 2, \ldots, N,$$

  donde E_i, $i = 1, 2, \ldots, N$ es un error aleatorio tal que:

  $$E_m(E_i) = 0 \qquad C_m(E_i, E_j) = \begin{cases} \sigma^2 v(x_i) & \text{si } i = j \\ \rho\sigma^2 \sqrt{v(x_i)}\sqrt{v(x_j)} & \text{si } i \neq j \end{cases}$$

  con β_0, β_1, $\sigma^2 > 0$ y ρ constantes desconocidas y $v(.)$ una función de x_i conocida con imagen en los reales positivos; también asumimos que $x_i > 0$, $i = 1, 2, \ldots, N$.

- **Modelo Clásico Estratificado** (C$_2(\beta_{0i}, \beta_{1i}; v(x_{ij}), \rho_i)$):

  $$Y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + E_{ij}, \quad j = 1, 2, \ldots, M_i, \quad i = 1, 2, \ldots, K,$$

  donde E_{ij}, $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, K$ es un error aleatorio tal que:

  $$E_m(E_{ij}) = 0 \qquad C_m(E_{ij}, E_{i^*j^*}) = \begin{cases} \sigma_i^2 v(x_{ij}) & \text{si } i = i^*,\ j = j^* \\ \rho_i \sigma_i^2 \sqrt{v(x_{ij})}\sqrt{v(x_{ij^*})} & \text{si } i = i^*,\ j \neq j^* \\ 0 & \text{si } i \neq i^* \end{cases},$$

  con β_{0i}, β_{1i}, $\sigma_i^2 > 0$ y ρ_i, $i = 1, 2, \ldots, K$, constantes desconocidas y $v(.)$ una función de x_{ij} conocida con imagen en los reales positivos; también asumimos que $x_{ij} > 0$, $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, K$.

4.1. Aportaciones población fija-modelo de superpoblación

4.1.1. *Estrategias óptimas que se apoyan en el modelo clásico base*

Cassel, Särndal and Wretman(1976,77) demuestran la optimalidad de la estrategia $(\overline{y}_s, \text{srswor})$ para el modelo $\text{C}_1(0, \mu; v(x_i) = 1, \rho)$ sin características auxiliares ($x_i = 1$, $i = 1, 2, \ldots, N$) entre los muestreos ppswor con probabilidades de inclusión estrictamente positivas y estimadores lineales e insesgados respecto al muestreo.

Para poblaciones finitas en las que disponemos de información auxiliar en forma de una covariable, Godambe and Joshi(1965) obtienen la optimalidad de la estrategia $(\widehat{\overline{y}}_{HT}, \text{ppswor})$ con $\pi_i = \frac{n\sqrt{v_i}}{\sum_{i=1}^{N}\sqrt{v_i}}$ para el modelo $\text{C}_1(0, \beta_1; v(x_i) = v_i, 0)$ con variables error independientes entre los muestreos ppswor con probabilidades de inclusión estrictamente positivas y estimadores de la forma $\widehat{\overline{y}} = \frac{n}{N}\overline{y}_s + \frac{N-n}{N}\widehat{\overline{y}}_u$ con $\widehat{\overline{y}}_u$ un estimador de $\overline{Y}_u$ insesgado respecto al procedimiento de muestreo. Debemos citar también los trabajos anteriores de Godambe(1955) y Hájek(1959) en los que se obtenía la optimalidad de esta misma estrategia para el modelo $\text{C}_1(0, \beta_1; v(x_i) = x_i^2, 0)$ minimizando $E_p V_m(\widehat{\overline{y}})$ entre los estimadores lineales y homegéneos insesgados respecto al procedimiento de muestreo (también sin reemplazamiento y con probabilidades de inclusión estrictamente positivas) y que fueran además insesgados respecto al modelo (se minimizaba así (sección 2,4) entre los estimadores que cumplían (sección 2,1) y (sección 2,3)). (Hájek(1959) obtiene este resultado como caso particular de un teorema en el que considerando un modelo más general que el modelo clásico base prueba que la estrategia $(\widehat{\overline{y}}_{HT}, \text{ppswor})$ con $\pi_i \propto \sqrt{V_m(Y_i)}/\sqrt{c_i}$, $i = 1, 2, \ldots, N$ minimiza $E_p V_m(\widehat{\overline{y}})$ entre los estimadores lineales y homegéneos insesgados respecto al procedimiento de muestreo y los procedimientos de muestreo sin reemplazamiento que cumplen que $C = \sum_{i=1}^{N} c_i \pi_i$.)

Si revisamos estos tres resultados observamos que la estrategia óptima de muestreo consiste en un muestreo ppswor con $\pi_i \propto \sqrt{V_m(Y_i)}$, $i = 1, 2, \ldots, N$. Este último resultado se obtiene también para los modelos de transformación e intercambiables estudiados por Cassel, Särndal and Wretman(1976,77), el modelo de regresión no independiente con coeficientes de regresión conocidos de Tam(1984), el modelo de Mukherjee and Sengupta(1989) y el modelo de permutación aleatoria de Rao(1975).

4.1.2. *Estrategias óptimas que se apoyan en el modelo clásico estratificado*

En poblaciones estratificadas, Cassel, Wretman and Särndal(1977) obtienen, para el modelo $\text{C}_2(0, \mu_i; v(x_{ij}) = 1, 0)$ sin características auxiliares ($x_{ij} = 1$, $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, K$), la optimalidad de la estrategia $(\widehat{\overline{y}}_{HT}, \text{ppswor})$ con $\pi_{ij} = Pr((i,j) \in s) = \frac{n\sigma_i}{\sum_{i=1}^{K} M_i\sigma_i}$ entre los muestreos ppswor con probabilidades de inclusión estrictamente positivas y estimadores lineales e insesgados respecto al muestreo.

Observemos que el muestreo stsrswor con asignación de Neyman (ver (sección 3,11)) satisface las condiciones anteriores, y que la estrategia de muestreo óptima vuelve a ser un muestreo ppswor con $\pi_{ij} \propto \sqrt{V_m(Y_{ij})}$, $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, K$. Se pueden obtener resultados similares respecto a la estrategia óptima de muestreo para modelos intercambiables (ver Thompson(1978)), modelos de transformación (ver Cassel, Särndal and Wretman(1976,77)) y para modelos de permutación aleatoria (ver Rao and Bellhouse(1978)).

Los modelos de superpoblación se han aplicado también cuando tenemos una característica auxiliar asociada a la característica en estudio para:

– Valorar la utilidad que puede obtenerse al aplicar esta información auxiliar en la construcción del estimador o en la estratificación de una población. Destacar en esta línea de investigación el trabajo de Särndal, Swensson and Wretman(1992).

– Valorar comparativamente entre dos grupos de estrategias en presencia de información auxiliar: las que introducen dicha información en la construcción del estimador pero no en el mecanismo de muestreo y las que la introducen en el estimador y también en el mecanismo de muestreo aplicado. En relación a esta línea de investigación podemos citar el análisis presentado en el manual de Cassel, Särndal and Wretman(1977) y en el artículo de Mukhopadhyay and Tracy(1993), y los artículos de Särndal(1980) y Padmawar(1981), entre otros.

4.2. Aportaciones predictivas

4.2.1. Estrategias predictivas basadas en el modelo clásico básico

Cassel, Särndal and Wretman(1977), comprueban la optimalidad de $\overline{y}_s$ para el modelo C$_1(0, \mu; v(x_i) = 1, \rho)$ sin características auxiliares ($x_i = 1$, $i = 1, 2, \ldots, N$) entre los estimadores lineales e insesgados.

Para poblaciones finitas en las que se dispone de información adicional en forma de una covariable, Brewer(1963) y Royall(1970) considerando el modelo C$_1(0, \beta_1; v(x_i), 0)$ con variables error independientes proponen el estimador:

$$\widehat{\overline{y}}_{b1} = \frac{n}{N}\overline{y}_s + \frac{N-n}{N}\widehat{\beta}_1 \overline{x}_u, \tag{1}$$

donde $\widehat{\beta}_1 = \frac{\sum_{i\in s} \frac{x_i y_i}{v(x_i)}}{\sum_{i\in s} \frac{x_i^2}{v(x_i)}}$, demostrando su optimalidad para el modelo anterior entre los estimadores lineales e insesgados. En este mismo trabajo, Royall(1970) propone la estrategia óptima de muestreo demostrando que si $v(x_i)$ y $\frac{v(x_i)}{x_i^2}$ son funciones no decrecientes la estrategia de muestreo óptima consiste en un muestreo intencionado en el que se seleccionan con probabilidad 1 los n índices distintos de la población con valor de la característica auxiliar más alto. En otro trabajo, Royall(1971) asumiendo el modelo C$_1(\beta_0, \beta_1; v(x_i) = 1, 0)$ con variables error independientes probó la optimalidad del estimador de regresión:

$$\widehat{\overline{y}}_{b2} = \overline{y}_s + \widehat{\beta}_2(\overline{x} - \overline{x}_s), \tag{2}$$

con $\widehat{\beta}_2 = \frac{\sum_{i \in s}(x_i - \overline{x}_s) y_i}{\sum_{i \in s}(x_i - \overline{x}_s)^2}$, en la clase de los estimadores lineales e insesgados. Demostrando además que la estrategia óptima de muestreo consiste en seleccionar una muestra balanceada de orden 1 (ver (3)).

El procedimiento de muestreo intencionado propuesto por Royall(1970) basa su optimalidad en el modelo C$_1(0, \beta_1; v(x_i), 0)$ con variables error independientes y en principio, si el vector poblacional no fuera realización de este modelo de superpoblación exactamente, los resultados podrían presentar sesgos considerables (críticas de Cox(1971) y Neyman(1971) a los trabajos de Royall(1970,71)). Como respuesta a esta crítica el artículo de Royall and Herson(1973a) estudia la robustez de los estimadores $\widehat{\overline{y}}_{b1}$ (expresión (1)) e $\widehat{\overline{y}}_{b2}$ (expresión (2)). Definiendo el concepto de muestra balanceada de orden L, como aquella muestra que cumple la ecuación:

$$\frac{\sum_{i \in s} x_i^l}{n} = \frac{\sum_{i=1}^N x_i^l}{N}, \; l = 1, 2, \ldots, L, \tag{3}$$

estos autores obtienen los siguientes resultados sobre la robustez de los estimadores de regresión propuestos:

(i) Si la muestra cumple (3) para $L = 1$, el estimador $\widehat{\overline{y}}_{b1}$ para $v(x_i) = x_i$ permanece insesgado y óptimo bajo los modelos: C$_1(\mu, 0; v(x_i) = 1, 0)$, C$_1(\beta_0, \beta_1; v(x_i) = 1, 0)$ y C$_1(\beta_0, \beta_1; v(x_i) = x_i, 0)$ todos ellos con variables error independientes.

(ii) Si la muestra cumple (3), el estimador $\widehat{\overline{y}}_{b1}$ para $v(x_i) = x_i$ permanece insesgado y óptimo para modelos de regresión polinómica de orden L (o inferior) con residuos E_i independientes tales que $V_m(E_i) = \sigma^2 p(x_i)$, $i = 1, 2, \ldots, N$ con $p(x)$ un polinomio de x de grado L (o inferior) que no puede contener potencias de x que no aparezcan en la regresión polinómica.

(iii) Si el procedimiento de muestreo (no necesariamente aleatorio) produce muestras balanceadas, el estimador $\widehat{\overline{y}}_{b1}$ es óptimo para los modelos indicados en (i), (ii) y para el modelo C$_1(0, \beta_1; v(x_i) = x_i, 0)$ con variables error independientes. Pero esta robustez tiene un coste de eficiencia (el mismo para cualquier $L \geq 1$) cuando el modelo correcto es $\mathrm{C}_1(0, \beta_1; v(x_i) = x_i, 0)$, debido a que el procedimiento de muestreo aplicado no es el óptimo (ver párrafos anteriores para ver el muestreo intencionado óptimo).

(iv) Si la muestra cumple (3), el estimador $\widehat{\overline{y}}_{b2}$ permanece insesgado y óptimo para todos los modelos de regresión polinomiales descritos en el epígrafe (ii), con la característica de que esta robustez no supone en este caso costes en eficiencia cuando el modelo correcto para la población es el modelo $\mathrm{C}_1(\beta_0, \beta_1; v(x_i) = 1, 0)$ con variables error independientes.

Debemos señalar que la obtención práctica de muestras balanceadas no es sencilla y que el cumplimiento de la relación (3) es en la práctica muy difícil de conseguir. Royall(1992) propone el uso de procedimientos aleatorios restringidos como protección ante muestras poco balanceadas (no como protección de una mala elección de modelo) que garanticen que la muestra sea balanceada. Un ejemplo de estos procedimientos de muestreo es el «basket method» de Wallenius(1980).

4.2.2. Una técnica predictiva de estratificación

Royall and Herson (1973b), considerando que la población está modelizada correctamente de acuerdo con el modelo $\mathrm{C}_1(0, \beta_1; v(x_i) = x_i, 0)$ con variables error independientes, proponen un procedimiento de estratificación y una estrategia conjunta de estimador y mecanismo de muestreo estratificado que son más eficientes que la estrategia de estimador $\widehat{\overline{y}}_{b1}$ con $v(x_i) = x_i$ y muestra balanceada (ver (3)) que comentábamos en la subsección 4.2.1. La propuesta de Royall and Herson también tiene la propiedad de robustez ante desviaciones polinómicas del modelo. Sin entrar en detalles sobre la forma del estimador, el procedimiento de estratificación y muestreo óptimo se obtienen a partir de las siguientes reglas:

(i) Los estratos se construyen en función de los valores de la característica auxiliar x_i, $i = 1, 2, \ldots, N$ ordenando los elementos de la población de menor a mayor valor de la covariable y formando los estratos con los M_1 primeros elementos, los M_2 elementos siguientes y así sucesivamente hasta formar el K-ésimo estrato con los M_K últimos elementos.

(ii) La muestra se obtiene mediante un muestreo estratificado (se muestrean todos los estratos) seleccionando (por un procedimiento de muestreo aleatorio o intencionado) en cada estrato una muestra balanceada de m_i elementos distintos.

(iii) Bajo el criterio de minimizar el error cuadrático del estimador respecto al modelo para un coste $C = c_0 + \sum_{i=1}^{K} c_i m_i$ dado, Royall and Herson(1973b) prueban que el número óptimo de elementos a seleccionar en cada estrato i, $i = 1, 2, \ldots, K$, viene dado por la siguiente expresión:

(4) $$m_i = n\frac{M_i\sqrt{\frac{\overline{x}_i}{c_i}}}{\sum_{i=1}^{K} M_i\sqrt{\frac{\overline{x}_i}{c_i}}}.$$

Observemos que esta expresión coincide cuando $c_i = c$, $i = 1, 2, \ldots, K$ con la asignación proporcional a la raíz de las medias en cada estrato (ver (sección 3,15)) que se propone en los trabajos desde la perspectiva de población fija para aproximar la asignación óptima de Newman.

4.2.3. *Estrategias predictivas basadas en el modelo clásico estratificado*

Royall(1976) obtiene para el modelo C$_2(0, \beta_{1i} = \mu; v(x_{ij}) = 1, \rho_i)$ sin características auxiliares ($x_{ij} = 1$, $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, K$) la expresión del estimador óptimo para $\overline{Y}$ entre los estimadores lineales e insesgados respecto al modelo para un muestreo en dos etapas sin reemplazamiento. En este mismo trabajo se realizan los siguientes comentarios sobre los tamaños de muestreo óptimos en un muestreo en dos etapas:

(i) Si $\sigma_i^2 = \sigma^2$ y $\rho_i = \rho > 0$ fijados k y n propone escoger en la primera etapa del muestreo las k unidades distintas de mayor tamaño.

(ii) Si $\sigma_i^2 = \sigma^2$ y $\rho_i = \rho = 0$ fijados k y n, el tamaño óptimo de muestreo de cada unidad seleccionada cumple que:

(5) $$m_i = \omega\frac{M_i n}{\sum_{i\in s_0} M_i} + (1-\omega)\frac{n}{k},$$

con $\omega = \frac{\sum_{i\in s_0} M_i}{N}$.

(iii) Si $\sigma_i^2 = \sigma^2$ y $\rho_i = \rho = 1$ fijados k y n, el tamaño óptimo de muestreo de cada unidad seleccionada es aquél que cumple (5) con $\omega = \frac{\sum_{i\in s_0} M_i}{\sum_{i\in s_0} M_i + k(N - \sum_{i\in s_0} M_i)}$.

(iv) Si $\sigma_i^2 = \sigma^2$, $\rho_i = \rho = 0$ ó 1 y $M_i = M$, $i = 1, 2, \ldots, K$, si tomamos $m_i = m$, $i \in s_0$, el tamaño óptimo de muestreo de cada unidad seleccionada es aquel que cumple para una función de costes $C(k, m) = c_1 k + c_2 n$ la siguiente expresión:

(6) $$m = \frac{n}{k} = \sqrt{\frac{c_1}{c_2}\frac{1-\rho}{\rho}}.$$

Observemos que esta expresión coincide con la obtenida por Hansen, Hurwitz and Madow(1953) para población fija (ver (sección 3,19)) sustituyendo R_y por ρ.

Observemos que los puntos (i), (ii) y (iii) solamente dan respuestas parciales al problema de tamaño óptimo porque no presentan propuestas para determinar el número óptimo de unidades a muestrear k, el punto (iv), en cambio, proporciona una respuesta a este problema más satisfactoria, ya que a partir de (6) y asumiendo un coste fijo C tendríamos que el número óptimo de unidades a muestrear sería $k = C/(c_1 + c_2\sqrt{[c_1(1-\rho)]/(c_2\rho)})$.

5. APORTACIONES AL DISEÑO DESDE LA PERSPECTIVA DE MODELOS DE SUPERPOBLACIÓN BAYESIANOS

Las aportaciones Bayesianas que presentaremos a continuación se basan en asumir un modelo de superpoblación Bayesiano, desde esta perspectiva la solución inferencial al problema de estimar $\overline{Y} = \frac{n}{N}\overline{y}_s + \frac{N-n}{N}\overline{Y}_u$ consiste en minimizar, dada una muestra s formada con n índices distintos, el error cuadrático del estimador $\widehat{\overline{y}} = \frac{n}{N}\overline{y}_s + \frac{N-n}{N}\widehat{\overline{y}}_u$ respecto a la distribución predictiva del vector $\mathbf{Y}_u|\mathbf{y}_s$. (En las expresiones anteriores representamos por $\widehat{\overline{y}}_u$ a un estimador de $\overline{Y}_u$ y por $\mathbf{Y}_u = (Y_i, i \notin s)$ y $\mathbf{y}_s = (y_i, i \in s)$ a los vectores poblacionales no observados y observados dada la muestra respectivamente.) De acuerdo con esta definición, el estimador Bayes se obtiene tomando $\widehat{\overline{y}}_u = E(\overline{Y}_u|\mathbf{y}_s)$ y el error cuadrático medio cometido viene dado por la expresión $V(\overline{Y}|\mathbf{y}_s) = \frac{(N-n)^2}{N^2}V(\overline{Y}_u|\mathbf{y}_s)$.

Los resultados que presentaremos a continuación se apoyan en los siguientes tres modelos Bayesianos:

– **Modelo Bayes Base** ($\mathrm{B}_1(\mu_0;\sigma^2,\sigma_0^2)$):

$$\begin{aligned} Y_i|\mu &\sim N(Y_i|\mu,\sigma^2), \text{ indep., } i = 1,2,\dots,N, \\ \mu &\sim N(\mu_0,\sigma_0^2), \end{aligned}$$

con $\sigma^2 > 0$, $\sigma_0^2 > 0$ y μ_0 escalares conocidos.

– **Modelo Bayes Estratificado** ($\mathrm{B}_2(\mu_0;\sigma_i^2,\sigma_0^2)$):

$$\begin{aligned} Y_{ij}|\mu_i &\sim N(Y_{ij}|\mu_i,\sigma_i^2), \text{ indep., } j = 1,2,\dots,M_i,\ i = 1,2,\dots,K, \\ \mu_i &\sim N(\mu_0,\sigma_0^2), \text{ iid, } i = 1,2,\dots,K, \end{aligned}$$

con $\sigma_i^2 > 0$, $i = 1,2,\dots,K$, $\sigma_0^2 > 0$ y μ_0 escalares conocidos.

– **Modelo Bayes en Dos Etapas** ($\mathrm{B}_3(\mu_0;\sigma_i^2,\sigma_\mu^2,\sigma_0^2)$):

$$\begin{aligned} Y_{ij}|\mu_i &\sim N(Y_{ij}|\mu_i,\sigma_i^2),\ \text{indep.},\ j=1,2,\ldots,M_i,\ i=1,2,\ldots,K,\\ \mu_i|\mu &\sim N(\mu_i|\mu,\sigma_\mu^2),\ \text{iid},\ i=1,2,\ldots,K,\\ \mu &\sim N(\mu_0,\sigma_0^2), \end{aligned}$$

con $\sigma_i^2>0$, $i=1,2,\ldots,K$, $\sigma_\mu^2>0$, $\sigma_0^2>0$ y μ_0 escalares conocidos.

Las notaciones indep. e iid significan repectivamente independientes e igualmente distribuidos, y las notaciones $\mathrm{B}_1(\mu_0;\sigma^2,\infty)$ y $\mathrm{B}_3(\mu_0;\sigma_i^2,\sigma_\mu^2,\infty)$ que aparecerán más adelante en el texto, representan la distribución inicial de referencia $\pi(\mu)\propto 1$.

A partir de las definiciones anteriores, observemos que:

– En el modelo $\mathrm{B}_1(\mu_0;\sigma^2,\sigma_0^2)$, las variables Y_i, $i=1,2,\ldots,N$ son intercambiables.

– En el modelo $\mathrm{B}_2(\mu_0;\sigma_i^2,\sigma_0^2)$, las variables Y_{ij}, $j=1,2,\ldots,M_i$ son intercambiables para $i=1,2,\ldots,K$.

– En el modelo $\mathrm{B}_3(\mu_0;\sigma_i^2,\sigma_\mu^2,\sigma_0^2)$, las variables $Y_{ij}|\mu_i$, $j=1,2,\ldots,M_i$ son intercambiables con medias por unidad, μ_i, $i=1,2,\ldots,K$ asimismo intercambiables.

La intercambiabilidad expresa, desde un punto de vista subjetivo, que no hay información sobre la característica en estudio y_i en el índice i, $i=1,2,\ldots,N$ y por lo tanto no hay razón para considerar que las inferencias basadas en una muestra de tamaño fijo n (con índices distintos) son mejores que las basadas en otra muestra distinta también del mismo tamaño (y también con índices distintos). Esta noción de intercambiabilidad ajusta con la idea objetiva de un muestreo srswor, y en consecuencia podemos asociar al modelo $\mathrm{B}_1(\mu_0;\sigma^2,\sigma_0^2)$ un muestreo srswor, al modelo $\mathrm{B}_2(\mu_0;\sigma_i^2,\sigma_0^2)$ un muestreo stsrswor y al modelo $\mathrm{B}_3(\mu_0;\sigma_i^2,\sigma_\mu^2,\sigma_0^2)$ un muestreo 2ssrswor.

Distinguiremos a continuación entre dos grupos de aportaciones: las relativas a la comparación desde una perspectiva Bayes de los muestreos srswor, stsrswor y clsrswor, y las relativas a los tamaños óptimos de muestreo.

5.1. Comparación Bayesiana entre los muestreos srswor, stsrswor y clsrswor

Bayarri and Font(1994,98) y Font(1995a) apoyándose en los modelos $\mathrm{B}_1(0;\sigma^2,\infty)$ para el modelo srswor y $\mathrm{B}_3(0;\lambda\sigma^2,(1-\lambda)\sigma^2,\infty)$ con $0<\lambda<1$ conocido para un muestreo 2ssrswor (que tiene como casos particulares para $k=K$ el muestreo stsrswor y para $m_i=M_i$, $i\in s_0$ el muestreo clsrswor) obtienen los siguientes resultados:

(i) Si $K = k$, $M_i = M$, $m_i = m$, $i = 1, 2, \ldots, K$, los estimadores Bayes para los modelos B$_1(0; \sigma^2, \infty)$ y B$_3(0; \lambda\sigma^2, (1-\lambda)\sigma^2, \infty)$ coinciden con la media muestral $\overline{y}_s$, y además:

$$V_{B_1(\text{srswor})}(\overline{Y}|\mathbf{y}_s) - V_{B_3(\text{stsrswor})}(\overline{Y}|\mathbf{y}_s) = \frac{N-n}{N}\frac{\sigma^2}{n}\rho. \tag{1}$$

(ii) Si $K < k$, $M_i = M$, $i = 1, 2, \ldots, K$ y $m_i = M$, $i \in s_0$, los estimadores Bayes para los modelos B$_1(0; \sigma^2, \infty)$ y B$_3(0; \lambda\sigma^2, (1-\lambda)\sigma^2, \infty)$ coinciden con la media muestral $\overline{y}_s$, y además:

$$V_{B_3(\text{clsrswor})}(\overline{Y}|\mathbf{y}_s) - V_{B_1(\text{srswor})}(\overline{Y}|\mathbf{y}_s) = \frac{N-n}{N}\frac{\sigma^2}{n}(M-1)\rho. \tag{2}$$

Donde $\rho = C(Y_{ij}, Y_{ij^*}|\mu)$, $j \neq j^*$, $i = 1, 2, \ldots, K$ para el modelo B$_3(0; \lambda\sigma^2, (1-\lambda)\sigma^2, \infty)$.

Observemos que (1) y (2) coinciden con los resultados aproximados desde la perspectiva de población fija (sección 3,4) y (sección 3,5) respectivamente, sustituyendo ρ por R_y y σ^2 por S_y^2 (la correlación y varianza del modelo por las poblacionales).

El artículo de Bayarri and Font(1998) estudia también la estimación de $\lambda = 1 - \rho$ para muestreos stsrswor y clsrswor, considerando el modelo B$_3(\mu_0; \lambda\sigma^2, (1-\lambda)\sigma^2, \sigma_0^2)$ y una distribución sobre λ uniforme (0,1).

5.2. Tamaños Bayes de muestreo óptimos

Murgui(1982) presenta para los modelos B$_1(\mu_0; \sigma^2, \sigma_0^2)$ y B$_2(\mu_0; \sigma_i^2, \sigma_0^2)$ los siguientes resultados sobre tamaño óptimo:

(i) Considerando el modelo B$_1(\mu_0; \sigma^2, \sigma_0^2)$, se obtiene bajo el criterio de minimizar el valor esperado respecto a un muestreo srswor de la función $P(n) = \gamma V(\overline{Y}|\mathbf{y}_s) + c_0 + c_1 n$ la siguiente expresión para el tamaño óptimo:

$$n = \frac{\sigma\sqrt{\gamma}}{\sqrt{c_1}}\left(1 + \frac{\sigma^2}{N\sigma_0^2}\right) - \frac{\sigma^2}{\sigma_0^2}, \tag{3}$$

que para $\sigma_0^2 \to \infty$ (distribución inicial no informativa) coincide con el tamaño óptimo desde la perspectiva de población fija para una muestra srswor (ver expresión (sección 3,8)) sustituyendo σ por S_y.

(ii) Considerando el modelo $\mathrm{B}_2(\mu_0; \sigma_i^2, \sigma_0^2)$, bajo la hipótesis de minimizar $E_{\text{stsrswor}} V(\overline{Y}|\mathbf{y}_s)$ para un coste $C = c_0 + \sum_{i=1}^K c_i m_i$ obtiene la siguiente expresión para los tamaños de muestreo óptimos para cada estrato:

$$m_i = \frac{(\sigma_i^2 + M_i\sigma_0^2)\sigma_i}{\sqrt{c_i}\sigma_0^2} \frac{C - c_0 + \sum_{i=1}^K \frac{\sigma_i^2 c_i}{\sigma_0^2}}{\sum_{i=1}^K \frac{\sigma_i^3\sqrt{c_i}}{\sigma_0^2} + \sum_{i=1}^K \sigma_i M_i \sqrt{c_i}} - \frac{\sigma_i^2}{\sigma_0^2}, \tag{4}$$

que para $\sigma_0^2 \to \infty$ (distribución inicial no informativa) coincide con los tamaños óptimos desde la perspectiva de población fija para un muestreo stsrswor (ver expresión (sección 3,9)) sustituyendo σ_i por S_{yi}, $i = 1, 2, \ldots, K$ y con la asignación de Neyman (ver expresión (sección 3,11)) cuando $c_i = c$, $i = 1, 2, \ldots, K$ con la misma sustitución.

Ericson(1969b) obtiene para un modelo estratificado general intercambiable una solución alternativa a (4) (exacta para algunos valores de C y un algoritmo iterativo para los restantes valores) para obtener el tamaño óptimo de los estratos que minimiza $E_p V(\overline{Y}|\mathbf{y}_s)$ bajo las restricciones: $\sum_{i=1}^K c_i m_i \leq C$ y $0 \leq m_i \leq M_i$, $i = 1, 2, \ldots, K$. Además, en un trabajo posterior, Ericson(1988), considerando un modelo en 2 etapas intercambiable más general que $\mathrm{B}_3(\mu_0; \sigma_i^2, \sigma_\mu^2, \sigma_0^2)$ y bajo la hipótesis de unidades del mismo tamaño ($M_i = M$, $i = 1, 2, \ldots, K$), resuelve el problema de minimizar $E_{\text{2ssrswor}} V(\overline{Y}|\mathbf{y}_s)$ bajo las restricciones: $c_1 k + c_2 n \leq C$, $1 \leq k \leq K$ y $0 \leq m_i \leq M_i$, $i \in s_0$. Traduciendo estos últimos resultados a nuestro modelo con $M_i = M$ y $\sigma_i^2 = \sigma^2$, $i = 1, 2, \ldots, K$, podemos escribir el siguiente algoritmo para determinar el número de unidades y tamaño de las unidades óptimos según este criterio:

Algoritmo (5)

(i) Si $\sigma_\mu^2 \leq \frac{\sigma^2}{M-1}$, tomar el entero l que cumpla $l \leq \frac{C}{c_1 + Mc_2} \leq l + 1$ y entonces:

$$\begin{cases} \text{si } C < (l+1)c_1 + (lM+1)c_2 & \text{tomar } k = l \\ \text{en otro caso} & \text{tomar } k = l + 1 \end{cases}$$

Además $m_i = M$, $i \in \{i_1, \ldots, i_l\}$ y m_{l+1} (tamaño de la unidad $l+1$ a muestrear si $k = l + 1$) es un entero tal que $m_{l+1} \leq \frac{C - (k+1)c_1 - kMc_2}{c_2} \leq m_{l+1} + 1$.

(ii) Si $\sigma_\mu^2 > \frac{\sigma^2}{M-1}$, encontrar k y m, con $m_i = m$, $i \in s_0$ que minimicen la siguiente expresión:

$$V^*(k, m) = \frac{(K-k)\sigma_\mu^2}{Kk} + \frac{(M-m)\sigma^2}{(M-1)mk}.$$

La expresión $V^*(k,m)$ tiene la misma forma que la expresión de la varianza para una muestra 2ssrswor desde la perspectiva de población fija, sustituyendo S^2_{yw} por $\frac{M}{M-1}\sigma^2 = E(\frac{\sum_{i=1}^K \sum_{j=1}^M (y_{ij}-\mu_i)^2}{N-K})$ y S^2_{yb} por $M\sigma^2_\mu = E(\frac{\sum_{i=1}^K M(\mu_i-\overline{\mu})^2}{K-1})$, (con $\overline{\mu} = \frac{\sum_{i=1}^K \mu_i}{K}$). Esta identificación entre las dos expresiones permite aplicar para minimizar $V^*(k,m)$ el algoritmo de Eisenhart (algoritmo (sección 3,18)) tomando $a = c_1 \frac{M}{M-1}\sigma^2$ y $b = c_2(\sigma^2_\mu - \frac{\sigma^2}{M-1})$. Como observación general a todos estos resultados debidos a Ericson, conviene recordar que aunque el criterio de estimador óptimo aplicado por Ericson(1969b,88) consiste en obtener el estimador lineal Bayesiano (el estimador lineal que minimiza el error cuadrático medio entre los estimadores lineales), éste coincide bajo la hipótesis de normalidad con el estimador Bayes (el aplicado por los restantes autores citados).

Murgui(1982) aporta también resultados de tamaño óptimo para un modelo Bayesiano normal-gamma sin características auxiliares, un modelo Bayesiano estratificado normal-gamma sin características auxiliares y una aproximación al tamaño óptimo para el modelo B$_3(\mu_0; \sigma^2_i, \sigma^2_\mu, \sigma^2_0)$.

Observemos también que para modelos de regresión lineal (cuando conocemos una característica auxiliar de la población), el procedimiento de muestreo óptimo que minimiza $E_p V(\overline{Y}|\mathbf{y}_s)$ y los objetivos de robustez respecto al modelo dan lugar a procedimientos de muestreo intencionados y a muestras balanceadas Bayesianas respectivamente (ver Murgui(1982)). En general, a partir de los resultados basados en modelos de superpoblación Bayesianos y considerando distribuciones iniciales de referencia, se recuperan los resultados clásicos predictivos sobre estimación y diseño.

6. CONCLUSIONES Y REFERENCIAS ADICIONALES SOBRE OTRAS APORTACIONES AL DISEÑO EN POBLACIONES FINITAS

Las metodologías basadas en población fija y en modelos de superpoblación clásicos y Bayesianos obtienen, en ausencia de características auxiliares, resultados de diseño muy «similares». Si revisamos esta exposición observamos que las únicas diferencias entre los resultados de población fija y de superpoblación clásicos surgen de tomar la varianzas respecto al modelo en vez de las varianzas poblacionales y entre los resultados de superpoblación clásicos y Bayesianos en la presencia en estos últimos de parámetros de la distribución inicial que modifican levemente los resultados. En cambio, la coincidencia entre los resultados basados en las dos aproximaciones al problema se «rompe» al introducir en la inferencia la información derivada de una característica auxiliar, defendiendo la primera un muestreo aleatorio y la segunda un muestreo intencionado.

Se ofrecen a continuación, precedidas por un título orientativo de su contenido, algunas referencias adicionales sobre aportaciones no analizadas en este trabajo en referencia al diseño en poblaciones finitas:

– **La solución minimax a los problemas de tamaño óptimo.** Consultar Solomon and Zacks(1970), Bellhouse(84), Chaudhuri(1988) y Bolfarine y Zacks(1991).

– **La aplicación del diseño experimental a los problemas de muestreo en poblaciones finitas.** Ver, por ejemplo, Rao(1979) que presenta un resumen con referencias sobre las aplicaciones más extendidas y dentro de las aportaciones a muestreo controlado, los artículos de Chakrabarti(1963), Srivastava and Collins(1985) y Srivastava and Ouyang(1992).

– **Aportaciones al diseño multiobjetivo.** Ver Malec(1995).

– **Obtención de estrategias óptimas asintóticas.** Ver Bellhouse(1984) y Bolfarine and Zacks(1992).

– **Aportaciones al diseño con más de una característica auxiliar.** Ver, por ejemplo, Rao(1993) en población fija, Royall and Pfeffermann(1982), Royall(1992b), Bolfarine and Zacks(1992) y Tam(1995) desde la perspectiva predictiva, y Draper and Guttman(1968b) desde la perspectiva Bayesiana para un muestreo en dos fases.

– **Procedimientos de muestreo en dos fases y secuenciales óptimos.** Consultar los manuales clásicos de población fija (por ejemplo Cochran(1977)) y desde un punto de vista Bayesiano, Zacks(1969) para la aproximación teórica al problema y DeGroot and Starr(1969) y Draper and Guttman(1968a,b) en un trabajo aplicado a la estimación de las medias poblacionales de una población finita.

– **Inferencias sobre la media poblacional a partir de una muestra obtenida a partir de una ruta aleatoria.** Ver: Kish(1965) desde la aproximación de población fija, y Font(1995a) y Bayarri and Font(1996) desde la aproximación Bayes.

7. AGRADECIMIENTOS

Quiero expresar aquí mi agradecimiento a la Dra. M.J. Bayarri por la lectura previa de este trabajo y sus muy útiles y adecuados comentarios. También deseo expresar mi gratitud al Dr. Carles M. Cuadras y a dos referees anónimos de la revista *Qüestiió* por sus comentarios y sugerencias. Este trabajo ha sido parcialmente financiado por el proyecto de la DGES PB96-0776.

8. REFERENCIAS

[1] **Bayarri, M.J.** and **Font, B.** (1994). «A (Bayesian) note on non-random samples from finite populations». In *1994 Proceedings of the section on Bayesian statistical science. ASA.*, 246-251.

[2] **Bayarri, M.J.** and **Font, B.** (1996). «Bayesian hierarchical models for random routes in finite populations». In *Data analysis and information system.* Bock and Polasek, eds. Springer. 301-312.

[3] **Bayarri, M.J.** and **Font, B.** (1998). «A Bayesian comparison of cluster, strata and random samples». *Journal of Statistical Planning and Inference*. Forthcoming.

[4] **Bellhouse, D.R.** (1984). «A review of optimal designs in survey sampling». *The Canadian Journal of Statistics*, **12**, 53-65.

[5] **Bolfarine, H.** and **Zacks, S.** (1991). «Bayes and minimax prediction in finite population». *J. Statist. planning and inference*, **28**, 139-151.

[6] **Bolfarine, H.** and **Zacks, S.** (1992). *Prediction Theory for finite populations*. Springer-Verlag. New York.

[7] **Brewer, K.R.W.** (1963). «Ratio estimation and finite population. Some results deducible from the assumption of an underlying stochastic process». *Aust. J. Statist.*, **5**, 93-105.

[8] **Cameron, J.M.** (1951). «Use of variance components in preparing schedules for the sampling of baled wool». *Biometrics*, **7**, 83-96.

[9] **Cassel, C.M, Särndal, C.E.** and **Wretman, J.H.** (1976). «Some results on generalized difference estimation and generalized regression estimation for finite populations». *Biometrika*, **63**, 615-620.

[10] **Cassel, C.M, Särndal, C.E.** and **Wretman, J.H.** (1977). *Foundations of inference in survey sampling*. John Wiley and Sons. New York.

[11] **Chakrabarti, M.C.** (1963). «On the use of incidence matrices in sampling from a finite population». *Journal of the Indian Statist. Assoc.*, **1**, 78-85.

[12] **Chaudhuri, A.** (1988). «Optimality of sampling strategies». *Handbook of Statistics*, **6**, 47-96.

[13] **Cochran, W.G.** (1977). *Sampling Techniques. Third Edition*. John Wiley and Sons, New York.

[14] **Cox, D.R.** (1971). «Discussion of Royall(1971)». *Foundations of statistical inference*. V.P. Godambe and D.A. Sprott, eds. Holt, Rinehart and Winston. Toronto. 275.

[15] **Dalenius, T.** and **Gurney, M.** (1951). «The problem of optimum stratification. II». *Skand. Akt.*, **34**, 133-148.

[16] **Dalenius, T.** and **Hodges, J.L.Jr.** (1959). «Minimum variance stratification». *Jour. Amer. Stat. Assoc.*, **54**, 88-101.

[17] **DeGroot, M.H.** and **Starr, N.** (1969). «Optimal two-stage stratified sampling». *The Annals of Math. Statist.*, **40**, 575-582.

[18] **Des Raj** (1954). «On sampling with probabilities proportionate to size». *Ganita*, **5**, 175-182.

[19] **Draper, N.R.** and **Guttman, I.** (1968a). «Some Bayesian stratified two-phase sampling results». *Biometrika*, **55**, 131-139.

[20] **Draper, N.R.** and **Guttman, I.** (1968b). «Bayesian stratified two-phase sampling results: k characteristics». *Biometrika,* **55**, 587-589.

[21] **Ekman, G.** (1959). «An approximation useful in univariate stratification». *Ann. Math. Stat.*, **30**, 219-229.

[22] **Ericson, W.A.** (1969a). «Subjetive Bayesian models in sampling finite populations». *J. Roy. Statist. Soc. (Ser. B)*, **31**, 195-233.

[23] **Ericson, W.A.** (1969b). «Subjetive Bayesian models in sampling finite populations. Stratification». In *New developments in survey sampling*. N.L. Johnson and Harry Smith Jr., eds. John Wiley and sons.

[24] **Ericson, W.A.** (1988). «Bayesian inference in finite populations». *Handbook of Statistics*, **6**, 213-246.

[25] **Font, B.** (1995a). *Análisis Bayesiano de muestras no aleatorias en poblaciones finitas*. Tesis doctoral. Universitat de València.

[26] **Font, B.** (1995b). «Inferencia Bayesiana en poblaciones finitas: un análisis comparativo». *Estadística Española*, **37, 120**, 309-320.

[27] **Godambe, V.P.** (1955). «A unified theory of sampling from finite population». *J. Roy. Statist. Soc. (Ser. B)*, **17**, 269-278.

[28] **Godambe, V.P.** (1992). «Survey Sampling - as I understand it». *Current issues in statistical inference. Essays in honor of D. Basu*. Ed. M. Ghosh and P.K. Pathak, **17**, 187-195.

[29] **Godambe, V.P.** and **Joshi, V.M.** (1965). «Admissibility and Bayes estimation in sampling finite population I». *Ann. Math. Statist.*, *36*, 1707-1722.

[30] **Hájek, J.** (1959). «Optimum strategy and other problems in probability sampling». *Časopis Pěst. Mat.*, **84**, 387-423.

[31] **Hansen, M.H., Hurwitz, W.N.** and **Madow, W.G.** (1953). *Sample survey methods and theory*. Vol. I and II. John Wiley and Sons. New York.

[32] **Hedayat, A.S.** and **Sinha, B.K.** (1991). *Design and inference in finite population sampling*. John Wiley and Sons. New York.

[33] **Joshi, V.M.** (1965). «Admissibility and Bayes estimation in sampling finite population II, III». *Ann. Math. Statist.*, **36**, 1723-1742.

[34] **Joshi, V.M.** (1967). «Confidence intervals for the mean of a finite population». *Ann. Math. Statist.*, **38**, 1180-1207.

[35] **Kish, L.** (1965). *Survey Sampling*. John Wiley and Sons. New York.

[36] **Konijn, H.S.** (1973). *Statistical theory of sample survey design and analysis*. North Holland. Amsterdam.

[37] **Kossack, C.F.** and **Shiledar Baxi, H.R.** (1971). «On designing of a unit-stratified survey design for discrete set of observations». *Internat. Statist. Rev.*, **39**, 46-56.

[38] **Lanke, J.** (1975). *Some contributions to the theory of survey sampling*. Ph. D. thesis. University of Lund.

[39] **Malec, D.** (1995). «Selecting multiple-objective fixed-cost sample designs using an admissibility criterion». *Journal of statistical planning and inference*, **48**, 229-240.

[40] **Mukherjee, R.** and **Sengupta, S.** (1989). «Optimal estimation of a finite population total under a general correlated model». *Biometrika*, **76**, 789-794.

[41] **Mukhopadhyay, P.** and **Tracy, D.S.** (1993). «Some inferential aspects in sampling finite populations under a fixed population set up: a review». *Journal of Applied Statistical Science*, **1**, 95-123.

[42] **Murgui, J.S.** (1982). *Diseño e inferencia en poblaciones finitas: modelos de superpoblación*. Tesis doctoral. Universitat de València.

[43] **Newman, J.** (1934). «On the two different aspects of the representative method. The method of stratified sampling and the method of purposive selection». *J. Roy. Stat. Soc.*, **97**, 558-625.

[44] **Neyman, J.** (1971). «Discussion of Royall(1971)». *Foundations of statistical inference*. V.P. Godambe and D.A. Sprott, eds. Holt, Rinehart and Winston. Toronto, 276-278.

[45] **Padmawar, V.R.** (1981). «A note on the comparison of certain sampling strategies». *J. Roy. Statist. Soc. (Ser. B)*, **43**, 321-326.

[46] **Rao, J.N.K.** (1975). «On the foundations of survey sampling». *A Survey of statistical design and linear models*. J.N. Srivastava, ed. North Holland, Amsterdam, 489-506.

[47] **Rao, J.N.K.** (1979). «Optimization in the desing of sample surveys». In J.S. Rustagi (ed.), *Optimizing Methods in Statistics: Proceedings of an International Conference*. New York. Academic Press. 419-434.

[48] **Rao, J.N.K.** and **Bellhouse, D.R.** (1978). «Optimal estimation of a finite population mean under generalized random permutation models». *Journal of statistical planning and inference*, **2**, 125-141.

[49] **Rao, T.J.** (1993). «On certain problems of sampling designs and estimation for multiple characteristics». *Sankhyā*, *Ser. B*, **55**, 372-384.

[50] **Reddy, V.N.** and **Rao, T.J.** (1977). «Modified PPS method of estimation». *Sankhyā*, *Ser. C*, **39**, 185-197.

[51] **Royall, R.M.** (1968). «An old appoach to finite population sampling theory». *J. Amer. Stat. Ass.*, **63**, 1269-1279.

[52] **Royall, R.M.** (1970). «On finite population sampling theory under certain linear regression models». *Biometrika*, **57**, 377-387.

[53] **Royall, R.M.** (1971). «Linear regression models in finite populations sampling theory». *Foundations of statistical inference*. V.P. Godambe and D.A. Sprott, eds. Holt, Rinehart and Winston. Toronto.

[54] **Royall, R.M.** (1976). «The linear least-squares prediction approach to two-stage sampling». *J. Amer. Stat. Ass.*, **71**, 657-664.

[55] **Royall, R.M.** (1982). «Balanced samples and robust Bayesian inference in finite population sampling». *Biometrika*, **69**, 401-409.

[56] **Royall, R.M.** (1988). «The prediction approach to sampling theory». *Handbook of Statistics*, **6**, 399-413.

[57] **Royall, R.M.** (1992a). «The model based (prediction) approach to finite population sampling theory». *Current issues in statistical inference. Essays in honor of D. Basu*. Ed. M. Ghosh and P.K. Pathak, **17**, 225-240.

[58] **Royall, R.M.** (1992b). «Robustness and optimal design under prediction models for finite populations». *Survey methodology*, **18**, 179-185.

[59] **Royall, R.M.** and **Herson, J.** (1973a). «Robust estimation in finite population I». *J. Amer. Stat. Ass.*, **68**, 880-889.

[60] **Royall, R.M.** and **Herson, J.** (1973b). «Robust estimation in finite population II: stratification on a size variable». *J. Amer. Stat. Ass.*, **68**, 890-893.

[61] **Särndal, C.E.** (1978). «Design-based and model-based inference in survey sampling». *Scand. J. Statist.*, **5**, 27-52.

[62] **Särndal, C.E.** (1980). «A two-way classification of regression estimation strategies in probability sampling». *The Canadian Journal of Statistics*, **8**, 165-177.

[63] **Särndal, C.E., Swensson, B.** and **Wretman, J.** (1992). *Model assisted survey sampling*. Springer-Verlag. New York.

[64] **Shiledar Baxi, H.R.** (1982). «Approximately optimum stratified design for a finite population». *Sankhyā*, *Ser. B*, **44**, 102-113.

[65] **Shiledar Baxi, H.R.** (1995). «Approximately optimum stratified design for a finite population II». *Sankhyā*, *Ser. B*, **57**, 391-404.

[66] **Solomon, H.** and **Zacks, S.** (1970). «Optimal design of sampling from finite populations: a critical review and indication of new research areas». *J. Amer. Stat. Ass.*, **65**, 653-677.

[67] **Som, R.K.** (1973). *A manual of sampling techniques*. Heinemann. London.

[68] **Srivastava, J.** and **Collins, F.** (1985). «On a general theory of sampling, using experimental design concepts I: estimation». *Bulletin de l'Institut International of Statistique*, **51, 10.3**, 1-16.

[69] **Srivastava, J.** and **Ouyang, Z.** (1992). «Sampling theory using experimental design concepts». In *Current issues in statistical inference: essays in honor of D. Basu*. M. Ghosh and P.K. Pathak, eds. 241-264.

[70] **Tam, S.M.** (1984). «Optimal estimation in survey sampling under a regression superpopulation model». *Biometrika*, **71**, 645-647.

[71] **Tam, S.M.** (1995). «Optimal and robust strategies for cluster sampling». *J. Amer. Stat. Ass.*, **90**, 379-382.

[72] **Thompson, M.E.** (1978). «Stratified sampling with exchangeable prior distribution». *The Annals of Statistics*, **6**, 1168-1169.

[73] **Tschuprow, A.A.** (1923). «On the mathematical expectation of the moments of frequency distributions in the case of correlated observation». *Metron*, **2**, 461-493, 646-683.

[74] **Wallenius, K.T.** (1980). «Statistical methods in sole source contract negotiation». *Journal of undergraduate mathematics and applications*, **0**, 35-47.

[75] **Yates, F.** (1960). *Sampling methods for censuses and surveys.* Charles Griffin and Co. London. Third Edition.

[76] **Zacks, S.** (1969). «Bayes sequential designs of fixed size samples from finite populations». *J. Amer. Stat. Ass.*, **64**, 1342-1349.

ENGLISH SUMMARY

A COMPARATIVE REVIEW OF DIFFERENT CONTRIBUTIONS IN SURVEY SAMPLING

B. FONT
Universitat de València*

Design is an important and applied area in finite populations. This article is a comparative review of some results in designing based on three different approaches: fixed population approach, inference based on classical superpopulation models and inference based on Bayesian superpopulation models. The article is organized in six sections, including an introduction. The contents of each section are described in this summary.

Keywords: Bayesian superpopulation models, classical superpopulation models, optimal design, sampling strategies, survey sampling.

AMS Classification (MSC 2000): 62D05.

---

* Begoña Font Belaire. Dept. Economia Financera i Matemàtica. Edifici Departamental Oriental. Av. dels Tarongers s/n. 46071 València (Espanya).

– Received August 1997.
– Accepted January 1999.

Section 2: *Basic concepts and notation*. We review the more basic concepts and definitions, and stablish the usual notation.

Section 3: *Contributions based on fixed population approach*. A review of the more known results in this approximation is presented including sampling designs, optimum size and stratification procedures. We point out the following results for later comparative reference in this summary:

(i) Earnings from stratification:

$$V_{\text{srswor}}(\overline{y}_s) - V_{\text{stratification}}(\overline{y}_s) \simeq \frac{N-n}{N}\frac{S_y^2}{n}R_y. \tag{1}$$

[Hansen et al(1953)]

(ii) Loses from cluster:

$$V_{\text{cluster}}(\overline{y}_s) - V_{\text{srswor}}(\overline{y}_s) \simeq \frac{N-n}{N}\frac{S_y^2}{n}(M-1)R_y. \tag{2}$$

[Cochran(1977)]

(iii) Random sampling optimum size:

$$n = \frac{S_y^2\sqrt{\gamma}}{\sqrt{c_1}}. \tag{3}$$

[Cochran(1977)]

(iv) Stratification optimum size:

$$m_i = n\frac{M_i S_{yi}}{\sum_{i=1}^{K} M_i S_{yi}}. \tag{4}$$

[Neyman(1934)]

(v) Two-stage sampling optimum size:

$$m_i = M_i\frac{K}{N}\sqrt{\frac{c_1}{c_2}\frac{1-R_y}{R_y}}, \tag{5}$$

[Hansen et al(1953)]

Section 4: *Contributions based on classical superpopulation models*. In this section we distinguish between contibutions based on design-unbiased estimation approach and on predictive approach. Point out:

(i) The agreement assuming a superpopulation model without auxiliary variable information with results in the fixed population approach. By example, Cassel et al(1977) with a design-unbiased estimation show the optimality of stratification with Neyman's assignment (see (4)). And Royall(1976) for the model:

$$Y_{ij} = \mu + E_{ij}, \qquad \mathrm{E}(E_{ij}) = 0, \quad \mathrm{C}(E_{ij}, E_{i^* j^*}) = \begin{cases} \sigma^2 & i = i^*, \quad j = j^* \\ 0 & \text{otherwise} \end{cases}$$

gets the following two-stage sampling optimum size with $M_i = M$ and $m_i = m$:

$$m = \frac{n}{k} = \sqrt{\frac{c_1}{c_2}\frac{1-\rho}{\rho}}. \tag{6}$$

Compare (6) with (5).

(ii) The disagreement assuming a superpopulation model with auxiliary variable information with the results based on the fixed population approach about optimum sampling design. So, Royall(1970, 71) and Royall and Heston(1973a,b) showed that the optimal sampling is a purposive design.

Section 5: *Contributions based on Bayesian superpopulation models*. We review two groups of contributions:

(i) The results of Bayarri and Font(1994) and Font(1995a) based on a Bayesian model without auxiliary variable information showing with $M_i = M$ and $m_i = m$ that:

$$V_{\text{srswor}}(\overline{Y}|\mathbf{y}_s) - V_{\text{stratification}}(\overline{Y}|\mathbf{y}_s) = \frac{N-n}{N}\frac{\sigma^2}{n}\rho, \tag{7}$$

and

$$V_{\text{cluster}}(\overline{Y}|\mathbf{y}_s) - V_{\text{srswor}}(\overline{Y}|\mathbf{y}_s) = \frac{N-n}{N}\frac{\sigma^2}{n}(M-1)\rho. \tag{8}$$

Compare (7) with (1), and (8) with (2).

(ii) The results about optimum size. Emphasize the results of Murgui(1982) and Ericson(1969b, 88). Point out the agreement assuming models without auxiliary variable information with optimum size results in fixed populations too, by example: Murgui(1982) shows for a random sampling plan that:

$$n = \frac{\sigma\sqrt{\gamma}}{\sqrt{c_1}}(1 + \frac{\sigma^2}{N\sigma_0^2}) - \frac{\sigma^2}{\sigma_0^2}, \tag{9}$$

and for a stratified sampling plan that:

$$(10) \qquad m_i = \frac{(\sigma_i^2 + M_i\sigma_0^2)\sigma_i}{\sqrt{c_i}\sigma_0^2}\frac{C - c_0 + \sum_{i=1}^K \frac{\sigma_i^2 c_i}{\sigma_0^2}}{\sum_{i=1}^K \frac{\sigma_i^3\sqrt{c_i}}{\sigma_0^2} + \sum_{i=1}^K \sigma_i M_i\sqrt{c_i}} - \frac{\sigma_i^2}{\sigma_0^2}.$$

Compare (9) with (3), and (10) with (4) for $\sigma_0^2 \to \infty$ (noninformative distribution) and $c_i = c$. And the disagreement assuming models with auxiliary variable information with the results based on fixed population approach.

Section 6: *Conclusions and additional references*. This section gives a main conclusion: «agreement between the three approachs when there aren't auxiliary variable information and disagreement in the otherwise», and additional references about other approximations to the design problems: minimax sampling designs, experimental designs, random routes, etc.