TRABAJO DE FIN DE GRADO

Estimación de Líneas de Pobreza con Información Auxiliar

Estimation poverty lines with auxiliary information

Autor: D./Dª Rosalía Romera Cantón.

Tutor/es: D./Dª Sergio Martínez Puertas.

Grado en Administración y Dirección de Empresas

Facultad de Ciencias Económicas y Empresariales

UNIVERSIDAD DE ALMERÍA

Curso Académico: 2013 / 2014

Almería, Junio de 2014

ÍNDICE

RESUMEN

El objetivo de este trabajo es tener una visión de los diferentes enfoques de pobreza (pobreza objetiva, pobreza subjetiva y pobreza multidimensional) basándose en el estudio de aquellas medidas de pobreza que vienen definidas como ratios de percentiles y centrándose en la estimación de este tipo de medidas cuando disponemos de una muestra extraída de una población finita mediante un determinado diseño muestral.

La incorporación de información auxiliar en el proceso de estimación de este tipo de medidas se llevará a cabo mediante la estimación eficiente de la función de distribución y de la estimación eficiente de cuantiles. De este modo, revisaremos las principales técnicas indirectas para la estimación de la función de distribución y revisaremos sus principales propiedades y a partir de ellas se definirán estimadores indirectos de los cuantiles que a su vez nos permitirán obtener estimaciones más eficientes de las medidas de pobreza.

Finalmente, se considerará un estudio de simulación con datos reales obtenidos de la Encuesta de Presupuestos Familiares correspondiente al año 2012y elaborada por el Instituto Nacional de Estadística (INE) .

ESTIMACIÓN DE LÍNEAS DE POBREZA CON INFORMACIÓN AUXILIAR

1. Introducción

Debido al gran interés que se tiene en la actualidad por los estudios en el análisis de la pobreza, el objetivo principal de este trabajo es poder dar una visión general sobre dicho análisis. Entre los motivos que propician este gran interés destacamos entre otros: que las líneas de pobreza nos indican el bienestar económico que tiene un país; sus problemas sociales…(European Commision, 1998; Kahn, 1998).

A continuación, se describen algunas de las técnicas utilizadas posteriormente durante el análisis para el estudio de las líneas de pobreza. En primer lugar, se analizan los diferentes enfoques de la pobreza que son, la Pobreza Objetiva (a ésta se le dará una mayor importancia), la Pobreza Subjetiva y la Privación Multidimensional.

Seguidamente, se muestran los métodos para la elaboración de las medidas de pobreza. Entre ellas podemos encontrar, el ratio del percentil 95/20, el ratio del percentil 95/50, el ratio del percentil 50/10, o los ratios 50/5 y 50/25. (Dickens y Manning, 2004).

Uno de los problemas que nos plantean estas medidas es la dificultad al incorporar información auxiliar, debido principalmente, a que dependen de los cuantiles, por lo que no son funciones lineales. Esto se puede mejorar con la incorporación de información auxiliar más eficiente cuando se determinen nuevos estimadores de cuantiles.

El empleo de las técnicas de calibración (Deville y Särndal, 1992; Rueda y otros, 2007) es el principal objetivo de este análisis, para poder obtener estimadores de la función de distribución, que incorporen información auxiliar eficiente. Con estos estimadores se pretende poder estimar de una manera eficiente el cuantil de cualquier orden, por lo que cualquier medida de pobreza que dependa de los cuantiles podrá ser estimada de manera eficiente.

En definitiva, el trabajo quedaría estructurado de la siguiente forma, en el apartado dos definiremos los diferentes enfoques de la pobreza centrándonos en la pobreza objetiva como ya mencionamos anteriormente; en el apartado tres hablaremos de las medidas de pobreza absolutas y relativas, centrándonos en esta ultima y comentando todas las posibilidades y disposición de las medidas. Por último, en el apartado cuatro haremos una estimación de las líneas de pobreza en poblaciones finitas con un diseño muestral, estimadores de la función de distribución y para cuantiles, terminando con un estudio de simulación con datos reales aportados por el INE.

Finalmente, decir que para que se pueda analizar la pobreza realizando todos los estudios anteriormente mencionados es imprescindible el conocimiento de la información estadística disponible como por ejemplo los datos y encuestas aportados por el INE.

2. Diferentes enfoques de la pobreza

El estudio del análisis de la pobreza resulta complejo debido al gran número de formas o enfoques para su medición o definición. Además, en este estudio, pueden influir muchos factores que pueden ser analizados de diversas maneras.

Como hemos dicho anteriormente, existen diferentes enfoques para analizar la pobreza, que dependen cada uno de ellos de la información a utilizar o de lo que se quiera resaltar, del punto de vista que queramos adoptar. Según la información base que se utilice se pueden distinguir entre Pobreza Objetiva y Pobreza Subjetiva. De la misma manera, pero utilizando otros parámetros podemos destacar a la Pobreza Relativa y a la Pobreza Absoluta.

Existen además, otros análisis para el estudio de los niveles de pobreza, los cuales no mencionamos por la no utilización de los mismo en este trabajo.

Y con una perspectiva totalmente distinta encontramos a la Privación Multidimensional. Éste análisis se encuentra basado principalmente en las limitaciones que pueden provocar la falta de integración social.

Los estudios de pobreza objetiva se caracterizan por ser el propio investigador el que determina y observa la información que se va a utilizar en dicho estudio, lo que le confiere un elevado grado de objetividad. Y la pobreza subjetiva, basa su análisis en la percepción que tienen los propios individuos sobre su situación.

2.1 Pobreza objetiva

Dentro de la pobreza objetiva se pueden realizar tanto los análisis de pobreza absoluta como relativa.

Respecto a la pobreza relativa, tema objeto de estudio, se dice que una persona es pobre cuando se encuentra en clara desventaja con respecto a su entorno tanto en temas económicos como sociales.

Esta clasificación no puede ser igual en todos los países, ya que por ejemplo en España una persona que gane 500 euros al mes estaría mal posicionada en la sociedad, mientras que esa misma persona con los mismos ingresos en algún país de África pertenecería a una clase rica.

Analizando la pobreza absoluta, definimos a ésta como la carencia de bienes y servicios básicos que tienen los individuos. Es decir, la falta de bienes como la alimentación, vivienda, ropa, etc. que soporta cada individuo.

Este concepto debería ser igual en todo el mundo, pero es muy complicado establecer indicadores puros de pobreza absoluta.

La pobreza es dinámica, ya que una persona que ahora es rica puede en un futuro no serlo y viceversa. Por lo que para el análisis de la misma se requieren estudios de periodos largos de tiempo en los que se tengan en cuenta los cambios sufridos por la misma. Estos son los llamados análisis de pobreza persistente o de larga duración.

La estadística europea persigue con esta clasificación de larga duración, no contar como pobres a individuos que atraviesan una situación transitoria de pobreza, sino que solamente estén considerados como pobres a las familias o individuos que presentan esta situación en el último año, o al menos, dos de los tres años anteriores.

Si se dispone de datos de los mismos individuos en diferentes periodos podemos estudiar la movilidad, observando las entradas y salidas en la pobreza. Con la movilidad, lo que se intenta obtener es si los pobres de un año son los mismos que en otro año, o si por el contrario los individuos pobres van cambiando. En los países en los que el estrato de pobres no cambia, hay que afrontar la situación de una manera diferente debido a la mayor gravedad que presentan con respecto a los países en los que si cambia, ya que en éstos existe una mayor movilidad, por lo que se puede pasar a ser pobre o dejar de serlo con más facilidad.

2.2 Pobreza subjetiva

Como ya he definido anteriormente, la pobreza subjetiva es la que determinan los propios individuos analizando la situación en la que se encuentran. Es decir, esta manera de ver la pobreza es la que dicen tener los propios individuos con respecto a su situación económica, frente a la pobreza objetiva que se determina con la observación y medidas de los investigadores.

2.3 Privación multidimensional

Este concepto se denomina privación debido a que tiene que ver con la exclusión social y la privación de bienes y servicios necesarios que pueden ser de primera necesidad o no. Los indicadores utilizados para medir este concepto suelen ser no monetarios o de privación. Esta gran variedad de indicadores nos van a proporcionar perspectivas diferentes de un mismo enfoque debido a las distintas maneras de medir y concebir la pobreza.

Para tener una mejor comprensión de la pobreza será necesario combinar todos los enfoques que nos dan una información diferente para de esta manera obtener un resultado más general.

A la hora de analizar la pobreza hay que destacar y tener en cuenta que cuando se realizan estos estudios no se están incluyendo en ellos a personas sin techo o que viven en alguna institución, por lo que la verdadera pobreza no consta en estos estudios.

3. Medidas de pobreza

Para medir la pobreza, en este trabajo vamos a utilizar las líneas de pobreza absoluta y relativa. Con estas líneas, se van a clasificar a los diferentes individuos como pobres o no pobres. Esto va a depender de la posición de la línea en la que se localicen cada uno de ellos. Para la medición de la pobreza se van a utilizar en estas líneas indicadores de valor monetario.

3.1 Líneas de pobreza absoluta

Todas las personas necesitamos de un mínimo de recursos para tener un bienestar básico, por lo qua la línea de pobreza absoluta nos mide el coste que supondría tener los recursos básicos necesarios para tener un mínimo bienestar.

Existen diferentes líneas de pobreza absoluta para medir si un ciudadano es pobre o no en cada una de ellas. Entre algunas de ellas podemos destacar la de Rowntree (1901). Esta línea se caracteriza por la compra de una cesta de bienes básicos para el consumo diario de una familia que junto con el alquiler de la casa y la gasolina para el coche supondrían el salario mínimo que debe tener una persona para no ser pobre. Es decir, que todo hogar que tenga unos ingresos inferiores a ese umbral será calificado como pobre.

Pero esta línea ha tenido desde su creación varios inconvenientes, y es que la gasolina y el alquiler no tienen por qué ser considerados como bienes básicos para que una persona pueda vivir bien, ya que eso dependerá de la sociedad en la que nos movamos.

Otra de las líneas que podemos mencionar en este apartado sería la que dota un salario al día para que una persona no sea pobre. Esto quiere decir que cualquier individuo que no disponga de ese dinero al día será calificado como pobre. Esta línea seria la que fija un dólar per cápita por día como necesario para cubrir los recursos mínimos considerados para que una persona no se califique como pobre.

Además, existen otras muchas más como la de Mollie Orshanski (1963, 1965) que es otra forma de medir la pobreza. Esta línea consiste en tomar el gasto en alimentación de los hogares o familias como una proporción del gasto total.

Pero esta línea también tiene problemas ya que en países con un desarrollo mayor, el porcentaje destinado a alimentación sobre el total será menor. Por lo que no sería una buena estimación de la pobreza absoluta.

Otra línea creada para valorar la pobreza de un individuo seria la que fija un porcentaje máximo dedicado a la alimentación de los hogares del total de ingresos recibidos en los mismos. De esta manera, todo aquel hogar o familia que gaste en alimentación un porcentaje mayor del total de sus ingresos será calificado como pobre.

Como hemos podido ver, las líneas de pobreza absoluta suelen presentar inconvenientes debido a la dificultad que presenta la creación de una de ellas para que se pueda utilizar en distintas sociedades y épocas.

Otra cosa a tener en cuenta, es que estas líneas de pobreza son más utilizadas y aceptadas en países subdesarrollados o menos desarrollados que en los países que cuentas con un mayor desarrollo.

3.2 Líneas de pobreza relativa

Al igual que en las otras líneas, las líneas de pobreza relativa lo que intentan es diferenciar a las personas consideradas como pobres por algún motivo de las no pobres.

A diferencia de las líneas absolutas en la que un aumento homogéneo del nivel de ingresos en la sociedad suponía una disminución de pobres, en las líneas de pobreza relativa un aumento homogéneo o de igual proporción de los ingresos en las familias de una sociedad proporcionaría las mismas tasas de pobreza que antes de aplicar el aumento.

Para que cambie el porcentaje de pobres calculado con estas líneas será necesario cambiar la distribución de la renta.

Las variables utilizadas para medir la pobreza por estas líneas relativas serian variables monetarias. Entre ellas podemos destacar a la variable de ingresos y a la de gastos. Con estas dos variables podemos decir si una persona es calificada como pobre o no pobre, es decir, se establece un porcentaje mínimo sobre cada una de las variables y sobre ese porcentaje se verá reflejada la pobreza del individuo. De esta manera, todo aquel individuo con un porcentaje menor al establecido será considerado como pobre, y el que supere dicho umbral será no pobre.

Debido a que esta línea para medir la pobreza relativa va a ser nuestro tema objeto de estudio nos vamos a centrar en ella y analizarla más detenidamente. Entre los procedimientos para medir la pobreza relativa destacamos:

- Elección de la variable monetaria
- Ingreso por unidad de consumo
- Escalas de equivalencias
- Fijación de las líneas de pobreza
- Incidencia, distribución e intensidad de la pobreza
- Otras medidas de pobreza
- Pobreza persistente o de larga duración

Desarrollando cada uno de ellos tenemos que lo normal en la ***elección de la variable*** es, como ya dijimos anteriormente, seleccionar aquella que represente el gasto o los ingresos del individuo. Pero estas dos variables presentan varios inconvenientes.

En el caso de los ingresos, estos pueden variar de un año a otro y no por ello tiene que cambiar el estilo de vida de sus hogares, ya que estos podían disponer de ahorros y de créditos… Además, esta variable no refleja los verdaderos ingresos que tiene una familia puesto que también dispone de otros bienes y recursos materiales en el hogar y fuera de él que no están reflejados en los ingresos pero que si aumentan el bienestar del hogar.

En cuanto al gasto, esta variable es más estable ya que una familia no disminuye puntualmente sus gastos por una pequeña disminución de los ingresos actuales.

Pero también presenta desventajas, ya que en muchas ocasiones los porcentajes de consumo tienen más que ver con el entorno que nos rodea y las costumbres adquirida a los largo de los años que de los recursos de los que dispone el hogar.

Debido a todas estas desviaciones obtenemos un resultado final que no se ajusta a la realidad, puesto que por ejemplo en los ingresos, los de los trabajadores por cuenta ajena nos daría unos resultados más ajustados a la realidad que los ingresos de los trabajadores por cuenta propia.

De la misma manera que en los ingresos los resultados obtenidos a través de los gastos en las familias al año también presentan desviaciones con respecto a la realidad, y es que estos datos están recogidos por encuestas realizadas en los hogares, los cuales no saben con exactitud la cantidad que han gastado a la semana, mes, o trimestre. Todas estas desviaciones o errores son inevitables, ya que provienen de las encuestas y por muy bien realizadas y diseñadas que estén siempre variaran un poco la realidad, debido a la gran dificultad que supone recopilar estos datos de una manera exacta.

Desde hace unos años la variable ingreso es la que se está utilizando en Europa para medir la pobreza y la exclusión social de los ciudadanos europeos, todo ello sabiendo que la elección de una variable monetaria para medir la pobreza con los datos expuestos no es nada fácil.

Analizando ahora el ***ingreso por unidad de consumo***, las líneas de pobreza basadas en esta variable se construyen de la siguiente forma:

En primer lugar hay que calcular el ingreso total de cada hogar. Este ingreso se calcula con las rentas de los trabajos por cuenta propia o ajena, rentas de capital, prestaciones sociales, pagos o devoluciones del IRPF, transferencias entre hogares, etc.

La decisión entre utilizar como variable de estudio a los hogares o a los individuos es una de las decisiones que afectan al resultado final del análisis. Últimamente se está utilizando al individuo como variable de estudio ya que es él al que le afecta la pobreza realmente, aunque se supone que la pobreza de las personas depende de los ingresos del hogar.

Para ver cómo influye el hogar en el individuo se le asigna a cada una de las personas de un hogar un ingreso. Este ingreso puede ser el ingreso per cápita o el ingreso por unidad de consumo. Como el propio nombre indica, el ingreso por unidad de consumo se refiere al ingreso total del hogar dividido entre el número de unidades de consumo, mientras que el per cápita, es el que se calcula dividiendo el ingreso total de un hogar entre el número de personas del mismo hogar.

De entre estos dos, se prefiere el ingreso por unidad consumo, ya que tiene en cuenta factores como las economías de escala.

En cuanto a la ***escala de equivalencias***, el objetivo principal es determinar un ingreso medio por persona del total que pertenece al hogar. Estas escalas intentan reflejar la realidad de los hogares ya que tienen en cuenta a las economías de escala y a las unidades de consumo.

Estas unidades de consumo equivalentes en los hogares nos dicen que el consumo de los niños en un hogar es diferente al consumo de los adultos, por lo que esta diferencia tiene que reflejarse en las unidades de consumo de dicho hogar.

Y en cuanto a las economías de escala en los hogares, provocan que un aumento de los individuos del hogar no tenga que conllevar un aumento proporcional del ingreso para que sigan teniendo el mismo bienestar.

Las escalas de equivalencia es lo que se utiliza para determinar las unidades de consumo. Entre las diferentes escalas a utilizar podemos mencionar las siguientes:

- ✓ *Escalas estadísticas*, que pueden ser escala de la organización para la cooperación y el desarrollo económica (OCDE) o escala de Oxford y escala de la OCDE modificada.
- ✓ *Escalas paramétricas*
- ✓ *Escala con dos parámetros*

Para la ***fijación de la línea de pobreza***, se requiere el cálculo de la mediana. Con ella, lo que se pretende es asignar un ingreso por unidad de consumo a cada individuo del hogar para determinar la pobreza. Lo que hay que hacer es ordenar de menor a mayor ingreso a cada uno de los individuos y se calcula el valor del ingreso por unidad de consumo que deja a su izquierda al 50% de los individuos. A la mediana se le pueden asignar vatios valores para determinar la pobreza. Si el valor asignado fuera por ejemplo 40, todos los individuos que se situaran a la izquierda de este valor serian pobres, ya que sus ingresos serian menores por unidad de consumo que el umbral asignado y los que estén a su derecha serian los no pobres.

Entre los valores o porcentajes que se pueden fijar para calcular la línea de pobreza estarían el 20 y 25 para estudiar la pobreza severa y el 40, 50, 60, o 70 por ciento. EUROESTAT fija actualmente este umbral en un 60% de la mediana de la distribución de los ingresos por unidad de consumo.

En la actualidad hay un creciente interés en los estudios sobre el análisis de la pobreza. Entre las razones por las cuales puede explicarse esta tendencia actual nos encontramos con el hecho de que las líneas de pobreza pueden ser un indicador del bienestar económico de un país, también pueden poner de manifiesto problemas de equidad social o de estratificación social.

Por estas razones, tales estudios se centran en el análisis de la distribución de ingresos, así como en el estudio de la desigualdad social, ya que esta última puede ser un factor de sustentación de diferencias generalizadas de los ingresos en países desarrollados (European Commision, 1998; Kahn, 1998).

Entre las medidas empleadas para la medición de la desigualdad salarial podemos encontrar el ratio del percentil 95 con el percentil 20, el ratio de percentiles 95/50, el ratio 50/10 o los ratios 50/5 y 50/25 analizados por Dickens y Manning (2004). De todas estas medias expuestas anteriormente, hay que señalar que en el presente trabajo se opta por usar como medida de pobreza el ratio 95/50.

Las medias de ***intensidad, distribución e incidencia*** son necesarias para abordar el estudio de la pobreza, por lo que las vamos a ver a continuación:

Con las ***medidas de incidencia*** lo que se intenta saber es a cuantas personas u hogares afecta el problema de la pobreza, es decir, la extensión que tiene en el total de la población o en un grupo concreto.

El indicador que mide la incidencia es el porcentaje de pobres, que se denomina tasa de pobreza o tasa de riesgo de pobreza (TP). Este porcentaje sobre las tasas de pobreza se puede calcular sobre la sociedad en su conjunto o sobre diferentes grupos dentro de una sociedad, como pueden ser: la edad, el nivel de estudios, el sexo, etc.

La fórmula para calcular dicha tasa es:

$$\text{Tasa\_pobreza(TP)}=P/n$$

Siendo la “p” el número de pobres y la “n” el total de personas tanto pobres como no pobres dentro del grupo que se esté analizando.

Las ***medidas de distribución***, como su propio nombre indica, nos proporcionan información acerca de cómo se encuentran distribuidos el total de pobres y las características que tienen cada uno de ellos.

Tomando como referencia a grupos de personas por edad, sexo, nivel de estudios, etc. Estas medidas nos indican el número de pobres que cabria en cada una de estas clasificaciones, pudiendo por tanto, asignar diferentes medidas para combatir la pobreza en cada una de ellas.

Y con las ***medidas de intensidad***, lo que se pretende averiguar es hasta qué punto la pobreza afecta a la población en su conjunto, es decir, el grado de pobreza que tienen los individuos considerados pobres.

Las medidas relativas de intensidad no son suficientes para medir la profundidad de la pobreza y las diferencias que existen entre estos ciudadanos y los que no son pobres, por lo que sería necesario utilizar también algún indicador de la profundidad de la pobreza, que junto con estas medidas relativas nos darán unos mejores resultados.

Estas nuevas medidas son:

- ✓ La brecha de pobreza (BP), que nos indica la distancia que hay de los pobres al umbral establecido para medir dicha pobreza.
- ✓ La brecha de ingreso (BI)
- ✓ Brecha relativa de pobreza

Con la utilización en su conjunto de estas medidas de incidencia e intensidad, se pueden obtener resultados muy detallados sobre la sociedad. Por ejemplo, se podrían dar sociedades en las que el número de pobres es elevado, pero se sitúan todos ellos cerca del umbral asignado en la mediana, pero también se podrían dar sociedades con un número menor de pobres que se encuentran más alejados de dicho umbral.

Pero para analizar con más profundidad la pobreza de un país, o tener en cuentas las características y desigualdades que existen entre los propios pobres se tienen en cuenta otras medidas.

A parte de las medidas expuestas anteriormente, existen ***otras medidas*** más complejas y difíciles de interpretar, debido a que en ellas contemplan características de las tres medidas mencionadas con anterioridad, es decir, proporcionan información sobre la incidencia, intensidad y desigualdad entre los pobres.

Estas nuevas medidas son:

- ✓ Índice de Sen, que es una suma ponderada de las brechas de pobreza de individuos.
- ✓ Índice de Thon, que es una variante del índice anterior.
- ✓ Familia de índices de pobreza de Foster, Greer y Thorbecke.
- ✓ Índice de Hagenaars

- ✓ Curvas TIP de pobreza. Estas curvas reflejan las tres dimensiones de la pobreza, incidencia, intensidad y desigualdad entre los pobres. Tienen una filosofía muy similar a la curva de Lorenz.

Y para terminar con las medidas relativas, vamos a incorporar a este análisis ***medidas de pobreza persistente o de larga duración***.

Estas medidas, llevan consigo la dimensión temporal, por lo que para calcular el número de pobres necesitan de información de varios años. El número de años serias 4 consecutivos en los que se le puede dar un umbral para determinar la pobreza distinta en cada año. Por tanto, analizando el número de pobres de cada año tendríamos que una persona o individuo será pobre si está calificada como tal el último año y al menos dos de los tres anteriores.

El objetivo de estas medidas temporales es no calificar como pobre a personas que hayan caído en la misma por un periodo corto de tiempo debido a circunstancias de un momento en concreto.

4. Estimación de Líneas de Pobreza en Poblaciones Finitas

El objetivo de este capítulo es obtener estimadores indirectos del ratio 95/50 como medida de pobreza seleccionada, es decir obtener estimadores capaces de incorporar de una manera eficiente la información auxiliar proporcionada por variables auxiliares. El principal problema que presentan todas las medidas anteriormente mencionadas, es que no son funciones lineales de los datos pues dependen de los cuantiles y esto hace muy difícil la incorporación de información auxiliar ya que los trabajos relacionados con la estimación de la mediana u otros cuantiles que incorporan información auxiliar son menos numerosos, que por ejemplo para el caso de la estimación de la media y por tanto un modo adecuado de mejorar la estimación de las medidas de pobreza, se puede alcanzar con la incorporación eficiente de la información auxiliar a la hora de definir nuevos estimadores de cuantiles. Para conseguir este objetivo, lo que nos proponemos en primer lugar, es la estimación de cuantiles a partir de la función de distribución, esto es, incorporar la información auxiliar en la estimación de dicha función para obtener nuevos estimadores más eficientes y a partir de estos últimos, mediante su inversión, obtener estimaciones de los cuantiles.

Así, en el primer apartado revisaremos los diseños muestrales que emplearemos a lo largo del trabajo, para después realizar una revisión de las principales técnicas indirectas existentes para la función de distribución en el apartado 2. Con el empleo de estas técnicas, podemos definir estimadores indirectos de los cuantiles, que nos permiten realizar estimaciones de la medida de pobreza considerada. Finalmente, emplearemos los estimadores obtenidos en un estudio de simulación, para analizar el comportamiento de dichos estimadores.

4.1 Diseño Muestral

Para el presente estudio, representamos con U a una población finita, que esta constituida por N unidades. El tamaño de U es N y los elementos de U se llamarán unidades de la población y se denotarán por u_1, u_2… cada una de estas unidades esta identificada por un número, por lo que tendremos las siguientes expresiones:

$$U = \{ u_1, u_2, \dots, u_N\}$$

$$U = \{1,2, \dots, N\}$$

Por esta razón, es por lo que a veces utilizaremos i en lugar de utilizar u_i para simplificar la notación.

Con el nombre de marco, encontramos a la lista que nos permite identificar cada una de las unidades de la población.

Las unidades poblacionales poseen muchas características de interés, algunas pueden ser conocidas y otras desconocidas. Supongamos, por ejemplo, que U está formada por los alumnos matriculados en una facultad. Al obtener la lista por ordenador de los alumnos, dichas unidades están en correspondencia uno a uno con el numero que tienen en dicha lista. A esta lista es a lo que llamamos marco de una población. Además, podemos saber varias características asociadas a ellos como la edad, el sexo, número de asignaturas aprobadas, etc. Sin embargo, existen otro tipo de características que serían desconocidas, como las asignaturas optativas que elegirá cada alumno para el próximo año, etc.

Representamos por la variable Y la característica de la población que deseamos estudiar, y que llamamos variable de estudio. El valor que toma dicha variable sobre la población es desconocido, pero vendrá dado por el valor que cada unidad u_i asigna a dicha característica, $Y_{i,}$

$$Y = \left\{Y_1,\ Y_{2,\dots,}Y_N\right\}$$

Dado que la variable Y que queremos estudiar representa la característica "no conocida" de una población U, para su conocimiento total solo hay un procedimiento, la investigación de todos los individuos de la población. Esta investigación exhaustiva recibe el nombre de censo.

No obstante, es frecuente que no deseemos "toda" esta información, sino alguna parte de ella o bien alguna función de los valores $\{Y_1, \dots, Y_N\}$, que denominamos "parámetros" y que representamos genéricamente por θ (Y). En general, esta función es una media, un total, un percentil, etc, y una inferencia sobre la misma puede realizarse al obtener la información de una parte de la población, que es lo que denominamos muestra.

La determinación de la muestra que permita conocer la información deseada sobre toda la población es un punto crucial de toda teoría estadística, pues como es lógico se desea que con el menor número de elementos en la muestra, lo que se conoce como tamaño muestral, y por ello con el menor número de elementos a estudiar, se obtenga un juicio aceptable sobre el verdadero valor de dicho parámetro, el cual se habría obtenido al realizar la investigación sobre toda la población. Si no se emplea en este paso correctamente el método científico podemos dar fácilmente resultados incorrectos y con muy poca precisión.

En cuanto al espacio muestral, nuestra intención es tomar subconjuntos en U para obtener la información que nos permita realizar buenas inferencias. A cualquier subconjunto de U le llamaremos muestra. El conjunto de todas las posibles muestras será el conjunto de todos los subconjuntos de U, que representaremos por Mu, y cuyo cardinal es 2^N.

Como en general Mu es un conjunto muy extenso, consideraremos en nuestros estudios subconjuntos él, $M \subseteq Mu$ que llamaremos espacio muestral, dependiendo del problema abordado. Los elementos de M se representan por m.

Entre las propiedades que caracterizan al conjunto M encontramos: tener todas sus muestras con el mismo número de elementos (espacio muestral de tamaño fijo), contener un cierto subconjunto de U en todas sus muestras (espacio muestral con elementos prefijados) y ser una participación de U (espacio muestral partición).

En cualquier caso, es deseable, y así se supondrá en el futuro, que toda unidad poblacional está en al menos una muestra.

El número de muestras que tiene el espacio muestral suele llamarse tamaño del soporte muestral, y lo representamos por n(M):

$$M = \{m_1 , m_2 , \dots , m_{n(M)}\}$$

El número de unidades de cada muestra m se denomina tamaño de la muestra o tamaño muestral y se representa por n(m), tenemos:

$$m = \{u_{i1} , u_{i2 ,\dots,} u_{in(m)}\}$$

Cuando el tamaño es fijo lo representamos simplemente por n.

Notemos que dependiendo del estudio a realizar, nuestro espacio muestral será diferente, desempeñando este un papel básico en la teoría de muestras finitas.

Escogido el espacio muestral, $M = \{m_1 , m_2 , \dots , m_{n(M)}\}$, hay que indicar el modo de elegir dichas muestras. Como antes indicábamos, esta elección no se hará de un modo caprichoso sino regida por las leyes del azar. Para ello basta definir una ley de probabilidad discreta sobre M,

$$P(\cdot) : M \longrightarrow [0,1]$$

tal que,

$$P(m) > 0 \ \forall m \in M$$

$$\sum_{m \in M} p(m) = 1$$

El par (M, p(·)) se denomina diseño muestral.

Ej: conocido también como muestreo aleatorio simple, lo detonaremos con la abreviatura MAS (N,n). En este diseño, el espacio muestral lo constituyen todas las muestras de Mu que tienen tamaño fijo n(m) = n. Este espacio muestral, M^n , tiene un tamaño del soporte igual a $\binom{N}{n}$. La distribución de probabilidad sobre dicho espacio muestral es uniforme, esto es,

$$P(m) = \frac{1}{\binom{N}{n}} \quad \forall m \in M_n$$

Como indicábamos anteriormente, toda unidad de la población debe de estar en alguna muestra, pero a veces deseamos que un individuo esté mas veces que otro en las diferentes muestras que forman el espacio muestral, o deseamos escoger la muestra a través de la elección de individuos, para ello es necesario conocer las denominadas probabilidades de inclusión.

Sea m ∈ M , para representar esta situación definimos la variable indicador de pertenencia a la muestra como una aplicación,

$$I_k : M \longrightarrow \{0,1\}$$

tal que,

$$\forall m \in M \; y \; k \in U \quad I_k(\mathrm{m}) = \begin{cases} 1 \; si \; u_k \in m \\ 0 \; en \; otro \; caso \end{cases}$$

I_k es una variable aleatoria definida sobre el diseño muestral (M, p(·)) y su distribución viene dada por :

$$\mathrm{P}\,\{I_k = 0\} = 1 - \mathrm{P}\{I_k = 1\}$$

$$\mathrm{P}\,\{I_k = 1\} = \sum_{\substack{m \in M \\ k \in m}} P(m) \triangleq \pi_k$$

Así, π_k es la probabilidad de que el elemento k esté en la muestra resultante del mencionado experimento aleatorio, y se denomina probabilidad de inclusión de primer orden. Como todo elemento, u_k, debe estar en el menos una muestra del diseño, ha de verificarse que $\pi_k > 0$, $\forall k \in U.$ Cuando el diseño verifica esta condición el muestreo correspondiente se denomina muestreo probabilístico.

Observemos que I_k es una variable aleatoria de Bernoulli, Be (π_k). Así pues, tendremos:

$$\mathrm{E}\,[I_k] = \pi_k$$

$$\mathrm{V}\,[I_k] = \pi_k(1 - \pi_k)$$

También tiene especial interés el indicador I_{kl} definido de modo análogo, pero sobre dos elementos muestrales:

$$I_{kl}(\mathrm{m}) = \begin{cases} 1 \; si \; u_k, u_l \in m \\ 0 \; en \; otro \; caso \end{cases}$$

Se verifica que $I_{kl} = I_k I_l$ aunque no son independientes. Nuevamente I_{kl} es una variable aleatoria de Bernoulli, Be (π_{kl}) , siendo:

$$\pi_{kl} = P\{I_{kl} = 1\} = \sum_{\substack{m \in M \\ k,l \in m}} P(m)$$

Las cantidades π_{kl} reciben el nombre de probabilidades de inclusión de segunda orden.

Observamos que:

$$\text{Cov}[I_k, I_l] = E\,[I_k, I_i] - E[I_k]E[I_l] = \pi_{kl} - \pi_k\,\pi_l$$

a esa cantidad se le representa por Δ_{kl} , y aparece frecuentemente en las expresiones de los estimadores y sus errores.

Un diseño deberá verificar, además de $\pi_k > 0, \forall k \in U$, la condición adicional,

$$\pi_{kl} > 0 \qquad \forall k \neq l \in U$$

En el presente trabajo vamos a emplear los siguientes diseños muestrales:

- ✓ Muestreo Aleatorio Simple

El muestreo aleatorio simple es el diseño muestral más estudiado, y dadas las propiedades que posee frente a las estimaciones de parámetros y de errores de muestreo, es un diseño muy utilizado.

Presenta el inconveniente de ser un diseño de tamaño fijo y ser fuertemente dependiente de disponer de un marco poblacional con sus elementos muy bien identificados, por lo que es frecuente tener dificultades en obtener una muestra que realmente pertenezca a dicho diseño.

En este diseño, el espacio muestral, M, está formado por todas las muestras de tamaño n fijo, y la distribución de probabilidad p(·) definida en M es la ley uniforme. Por ello, puede definirse como,

$$P: Mu \longrightarrow [0,1] \text{ tal que } p(m) = \begin{cases} \frac{1}{\binom{N}{n}}\; si\; n(m) = n \\ 0\; si\; n(m) \neq n \end{cases}$$

Los parámetros asociados a este diseño muestral son:

$$\pi_i = \frac{n}{N} = f \qquad i = 1, \ldots, N$$

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)} \qquad i \neq j = 1, \ldots, N$$

Por lo que es un diseño muestral cuantificable. Y dado que es de tamaño fijo, y que,

$$\Delta_{ij} = \pi_{ij} - \pi_i \, \pi_j = -\frac{f\,(1-f)}{N-1} \qquad I \neq j$$

$$\Delta_{ii} = f\,(1\text{-}f)$$

Podemos fácilmente tener estimadores no negativos del error.

- ✓ Muestreo Sistemático

Hasta ahora hemos considerado una población U = $\{u_1, \ldots, u_N\}$ en la cual sus unidades están perfectamente identificadas. Imaginemos la población formada por los asistentes a un teatro, si deseamos hacer una encuesta sobre las bondades de la representación, ciertamente tenemos numerados los individuos que componen la misma por su billete de entrada, pero al realizar la elección de la muestra en base a él puede ser difícil escoger los individuos que nos da el mecanismo de azar. Además los individuos son ese día una representación de los diferentes individuos que pueden acceder a la función a lo largo del tiempo que la obra esté en cartel.

En estos casos se suele tomar la muestra usando el denominado muestreo sistemático que trata de escoger los individuos de la población, para pertenecer a la muestra, de un modo directo, seleccionando las unidades mediante una regla sistemática, partiendo de una elección inicial o primaria. Esta regla sistemática nos da siempre la misma muestra cuando se parte de la misma unidad primaria por ello las muestras pertenecientes a estos diseños son excluyentes entre sí.

Así, a la salida del teatro podemos fijar la regla de preguntar a una de cada diez personas, partiendo del tercer individuo que salga inicialmente. Con ello podemos escoger muestras de la población.

Observemos, no obstante, que además de los inconvenientes que presenta el que las muestras sean excluyentes, en el plano teórico es importante la regla de elección de la muestra, ya que si la población presenta algún tipo de tendencia puede ocurrir que la muestra sistemática no sea capaz de recogerla. Por tanto, cuando la población está distribuida al azar estas reglas sistemáticas pueden dar resultados similares a los que da el muestreo aleatorio simple, siendo más fácil la elección de la muestra.

Un diseño muestral sistemático es un diseño muestral cuyo espacio muestral está formado por muestras sistemáticas, es decir, las unidades se escogen de la población mediante la elección de una unidad de partida y las demás por una regla sistemática de selección. De este modo, seleccionamos una unidad al azar γ y a partir de ella se seleccionan sistemáticamente las unidades siguientes mediante un paso K, esto es las unidades seleccionadas son γ; $\gamma+K$, ... hasta agotar la población, suponiendo $N=Kn+r$ con n el tamaño muestral deseado y $0\leq r\leq K$

Las probabilidades de inclusión son fáciles de calcular, y vienen dadas por:

- De primer orden.

$$\pi_i = \frac{1}{K} \qquad \forall i \in U$$

luego es un muestreo probabilístico.

- De segundo orden.

$$\pi_{ij} = \begin{cases} \frac{1}{K} & si\ i, j\ pertenecen\ a\ la\ misma\ muestra \\ 0 & en\ otro\ caso \end{cases}$$

Al no ser $\pi_{ij} > 0,\ \forall i, j,$ no permite una estimación inmediata del error de muestreo, es decir, no es un muestreo cuantificable.

Otro inconveniente que presenta este diseño es el poco control que se tiene del tamaño muestral. En defecto, ya hemos visto que dado un valor de K, el tamaño de la muestra puede ser n ó n + 1.

- ✓ Muestreo Estratificado

La ciencia utiliza la homogeneización por bloques para obtener mayor precisión en sus estudios, pues de este modo puede controlar ciertos aspectos del experimento, para conocer la influencia de los mismos en los resultados.

Análogamente, se puede realizar en la Teoría del Muestreo, pues si muestreamos en una población muy heterogénea, se necesita un gran esfuerzo muestral para obtener cierta precisión, mientras que si la población está dividida en grupos internamente homogéneos, el esfuerzo en cada grupo será mínimo, resultando globalmente un esfuerzo menor.

Así por ejemplo, si deseamos obtener una muestra en determinada Facultad para conocer la media de horas de estudio que los alumnos han realizado en la pasada semana, será menos costoso si se efectúa por curso, ya que los alumnos de cada curso tienen similar problemática (exámenes, horarios de clases, etc) por lo que la muestra será pequeña en cada curso. Pero si deseamos cierta precisión y muestreamos globalmente, la muestra será de gran tamaño para que pueda recoger todas las tendencias que existen en la Facultad.

Además de las ventajas de esta homogeneización para reducir el error de muestreo, ya que a un mismo esfuerzo muestral debe corresponder mayor precisión, existen otras ventajas secundarias que aconsejan la división de la población en estratos,

- Facilidad de manejo de los marcos de los estratos.
- Procedimientos muestrales más ágiles en cada estrato.
- Mejor empleo de informaciones especiales.

En principio es fácil definir una población estratificada, y establecer el espacio muestral y las probabilidades asociadas, a partir de los diseños muestrales en cada estrato. Posteriormente veremos que esta división de la población en estratos deberá realizarse con algún patrón de homogeneidad para que dicha división aporte alguna información útil, que produzca una disminución en el error de las estimaciones.

Una población U se dice que esta estratificada en L estratos, $U_1, U_2, \ldots, U_L$, si estos forman una partición de la misma, es decir,

- $U = \bigcup_{i=1}^{L} U_i$
- $U_i \cap U_j = \emptyset, \quad \forall i \neq j$

Llamaremos N_h al número de individuos del estrato U_h, y supondremos $N_h \geq 2$ para evitar casos triviales. Se tiene pues $\sum_h N_h$ = N. a los elementos de la población que están en el estrato U_h los representamos por:

$$u_{h1}, u_{h2}, \ldots, u_{hNh}$$

y al cociente N_h / N se le llama peso del estrato U_h, representándose por W_h.

Una vez dividida la población en estratos, no se define sobre ella un diseño muestral directamente, sino de un modo indirecto, pues las muestras se escogen de modo independiente en cada uno de los estratos, agrupándose posteriormente para producir las muestras sobre U, siendo esta posibilidad de muestrear en cada estrato la gran ventaja de la división de la población.

Por ello, se define en cada estrato U_h un diseño muestral $d_h = (M_h, p_h(\cdot))$, independientemente de los demás estratos, obteniéndose en él la muestra m_h con probabilidad $p_h(m_h)$, estando m_h compuesta por elementos de U_h únicamente.

El espacio muestral M de U es el dado por M_1 x … x M_h, entendiendo que una muestra m ∈ M está formada por los individuos de las muestras $m_1, m_2, \dots, m_L$, que la definen, esto es, m = $(m_1, m_2, \dots, m_L)$, ya que las unidades no se repiten, con $m_h \in M_h$. y debido a la independencia entre los diseños muestrales en cada estrato, se verifica:

$$P(m) = p_1(m_1)\, p_2(m_2) \dots p_L(m_L)$$

Lo que define el diseño muestral sobre U, (M,p(·)), de forma indirecta, a partir de los diseños de cada uno de los estratos, estando la elección de estos diseños en función de la naturaleza de la información disponible en cada estrato.

Si el tamaño de la muestra m_h es n_h, el tamaño de la muestra m será $n = \sum_h n_h$. Al cociente n_h / N_h se le denomina fracción de muestreo o tasa de muestreo del estrato U_h, y se le representa por f_h.

Observemos que π_i, la probabilidad de que el individuo u_i esté en la muestra m, es igual a $\pi_i^{d_h}$, probabilidad de inclusión en la muestra m_h, si el elemento u_i está en el estrato U_h.

Análogamente, π_{ij} es igual a $\pi_{i_j}^{d_h}$, si ambos elementos están en U_h. Pero si están en estratos diferentes, U_h y U_k, entonces:

$$\pi_{ij} = \pi_i^{d_h}\ \pi_j^{d_k}$$

debido a la independencia que existe en la elección de la muestra.

Cuando no exista posibilidad de confusión, nos referiremos a la matriz del diseño $\pi = (\pi_{ij})$ sin utilizar superíndices.

Por consiguiente, definida la estratificación de la población y el diseño muestral en cada estrato, tenemos establecido el diseño muestral de la población completa, y trabajamos con él en el modo usual para realizar la estimación de parámetros, aunque como veremos, nos conduce a una mixtura de las estimaciones que podríamos realizar sobre los parámetros en cada estrato, como si fuera una población independiente. Asimismo obtendremos estimación de los errores globales a partir de los errores en los estratos.

En el caso de que en cada estrato se realice un muestreo MAS (N_h, n_h), se dice que el diseño muestral es aleatorio simple estratificado, y se denota MASE (N, n, L, $\{N_h, n_h\}_{h=1,\dots,L}$). Este diseño es uno de los más empleados pues la estratificación suple a veces el empleo de probabilidades variables, por ello vamos a calcular los estimadores y sus errores.

Para el estrato U_h tendremos:

$$\pi_i^h = \frac{n_h}{N_h}, \qquad \pi_{ij}^h \frac{n_h(n_h - 1)}{N_h\,(N_h - 1)}$$

Apartir de ahora, cuando hablemos de muestreo estratificado, se entenderá que nos estamos refiriendo al muestreo estratificado aleatorio simple.

4.2 Estimadores de la función de distribución

Sea U una población finita formada por N diferentes unidades, y sean Y_1, Y_2,…Y_N los valores que una variable de estudio y toma en los distintos elementos de U.

Sea $X_i = (X_{i1},\dots X_{ij}, \dots X_{ij})$ el valor de un vector de variables auxiliares en la unidad i de la población. Supongamos que una muestra s, es extraida de la población U de acuerdo con un determinado diseño muestral, cuyas probabilidades de inclusión de primer orden vienen dadas por $\pi_i = P[\ i \in s]$. Además se supone que el valor y_i de la variable de estudio y sólo está disponible para las unidades i incluidas en la muestra s, mientras que el valor x_i del vector auxiliar es conocido para todos los elementos de la población U.

La función de distribución poblacional de la variable y viene dada por:

$$F_y(t) = \frac{1}{N} \sum_{i \in U} \Delta(\mathrm{t} - y_i) \qquad (1, 4.2)$$

donde

$$\Delta(t - yi) = \begin{cases} 1 \ si \ t \geq y_i \\ 0 \ si \ t < y_i \end{cases}$$

El estimador usual de F_y (t), es el estimador de Horvitz-Thompson, el cual vienen dado por:

$$\hat{F}_{YH}(t) = \frac{1}{N}\sum_{i\in s}\Delta(t - y_i) \qquad (2, 4.2)$$

donde $d_i = \frac{1}{\pi_i}$

Entre las ventajas del estimador de Horvitz-Thompson tenemos que es insesgado.

Entre los problemas de este estimador esta que no usa o incorpora información auxiliar, por lo que usaremos de otras variables que nos den tal información.

Para incorporar información auxiliar existen varias alternativas:

- ✓ Estimadores de razón:

Los estimadores de razón se pueden construir considerando como variable de estudio, Δ (t-y_k), y como variable auxiliar, $\Delta(t - \hat{R} \cdot X_k)$, y siendo $\hat{R}$, el estimador usual de la razón de las medias (o totales) de las variables x e y, es decir:

$$\hat{R} = \frac{\sum_{k\in s} d_k y_k}{\sum_{k\in s} d_k x_k} \qquad (3, 4.2)$$

obteniendo así el estimador,

$$\hat{F}_R(t) = \frac{1}{N}\frac{\sum_{k\in s} d_k \Delta(t-y_k)}{\sum_{k\in s} d_k \Delta(t-\hat{R}\cdot x_k)} \cdot \sum_{k\in U}\Delta(t - \hat{R}\cdot x_k) \qquad (4, 4.2)$$

Observamos que este estimador, se basa implícitamente, en la existencia de una relación de proporcionalidad, más o menos aproximada, entre las variables x e y. En caso de ser y exactamente proporcional a x, se tendrá $\hat{F}_R$ (t) = F (t), lo que justifica la construcción de dicho estimador. De cualquier forma, la correlación entre $\Delta(t - \hat{R}\cdot x_k)$ y $\Delta(t - y_k)$, es usualmente más débil, que la existencia entre x e y, por lo que la ganancia de eficiencia de $\hat{F}_R$ (t) sobre $\hat{F}_{YH}$ (t), será menor que la obtenida cuando se emplea estimadores de razón para estimar $\overline{Y}$.

- ✓ Estimadores de diferencia:

También es posible construir estimadores diferencia para la función de distribución, volviendo a considerar como variable de estudio, Δ (t - y_k) y como variable auxiliar,

$\Delta(t - \hat{R} \cdot x_k)$, donde $\hat{R}$ ha sido definido previamente cuando se introdujo el estimador razón, obteniendo así el siguiente estimador:

$$\hat{F}_{\mathrm{d}}(\mathrm{t}) = \frac{1}{N} \sum_{k \in s} d_k \, \Delta(t - y_k) + d \left[\sum_{k=1}^{N} \Delta\left(t - \hat{R} \; x_k\right) - \sum_{k \in s} d_k \Delta(t - \hat{R} \, x_k) \right] \qquad (5{,}4.2)$$

donde d, es una constante, arbitrariamente elegida. El estimador $\hat{F}_{\mathrm{d}}$ (t) es asintóticamente insesgado y el valor óptimo de d es obtenido minimizando la varianza asintótica de $\hat{F}_{\mathrm{d}}$ (t), con respecto a d. Esta varianza viene dada por:

$$\mathrm{V}\left(\hat{F}_{\mathrm{d}}(\mathrm{t})\right) = \frac{1}{N^2} \; \widetilde{V} \; [\Delta(t - y_k) - \Delta(t - R x_k)] \qquad (6,\ 4.2)$$

De este modo, el valor óptimo de d, bajo muestreo aleatorio simple, viene dado por:

$$d^* = \frac{\rho_\Delta \cdot S_{\Delta Y}}{S_{\Delta X}}$$

donde

$$S_{\Delta Y} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\Delta(t - y_k) - F_y(t)\right)^2 \qquad (7,\ 4.2)$$

$$S_{\Delta X} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\Delta\left(t - \hat{R} \, x_k\right) - \frac{1}{N} \sum_{k=1}^{N} \Delta(t - \hat{R} \, x_k)\right)^2 \qquad (8,\ 4.2)$$

y donde ρ_Δ, es el coeficiente de correlación, en toda la población, entre $\Delta(t - yk)$ y $\Delta(t - \hat{R}\,\mathrm{xk})$.

- ✓ Estimador de Chambers y Dunstan:

Este estimador se basa en la existencia de un modelo de superpoblación, que liga la variable de estudio con las variables auxiliares, una o varias. Por supuesto, la eficiencia de este enfoque, dependerá de la exactitud del modelo que se considere. Supongamos que el modelo de superpoblación, es de la forma:

$$\mathrm{Y_k} = \alpha + \beta x_{\mathrm{k}} + \in_{\mathrm{k}}, \quad \mathrm{E}\varepsilon[\epsilon_k] = 0 \quad \mathrm{V}\varepsilon[\epsilon_k] = \sigma_\epsilon^2 \qquad (9,\ 4.2)$$

donde ϵ_{k} son errores aleatorios independientes e idénticamente distribuidos, y donde α y β, son parámetros superpoblacionales desconocidos.

Chambers y Dunstan (1986) construyen un estimador para los residuos. Para ello, es fácil ver:

$$\mathrm{P}[y_k \leq t] = \mathrm{G}\,[t - \alpha - \beta x_k] \qquad (10, 4.2)$$

donde G (t), es la función de distribución común de los errores ϵ_k. Se tiene entonces, para la esperanza en el modelo,

$$\mathrm{E}\varepsilon\,[\,\textstyle\sum_{k\in U-s} \Delta\,(t - y_k)] = \sum_{k\in U-s} P[y_k \leq t] = \sum_{k\in U-s} G(t - \alpha - \beta x_k) \qquad (11, 4.2)$$

siendo un estimador natural de $\mathrm{E}\varepsilon[F_c(t)]$,

$$\hat{F}_r\,(t) = \frac{1}{N-n} \sum_{k\in U-s} \hat{G}_k(t) \qquad (12, 4.2)$$

donde

$$\hat{G}_k(t) = \sum_{j\in s} \Delta\,(t - \hat{\alpha} - \hat{\beta}\; x_k - r_j)$$

y $r_j = y_j - \hat{\alpha} - \hat{\beta} x_j$, obteniéndose así, el siguiente estimador,

$$\hat{F}_{\mathrm{CD}}(\mathrm{t}) = \frac{n}{N}\, F_{\mathrm{m}}\,(\mathrm{t}) + \frac{N-n}{N}\, \hat{F}_r(t) \qquad (13, 4.2)$$

que es un estimador asintóticamente insesgado, bajo el modelo.

- ✓ Estimador de Rao, Kovar y Mantel:

Rao, Kovar y Mantel (1990), propusieron un estimador basado en el diseño, el cual es tanto asintóticamente insesgado bajo el diseño, como insesgado bajo el modelo. Dicho estimador, viene dado por:

$$\hat{F}_{\mathrm{RKM}}\,(\mathrm{t}) = \frac{1}{N}\left[\sum_{k\in s} d_k \Delta(t - y_k) + \sum_{j=1}^{N} \hat{G}_j - \sum_{k\in s} d_k \hat{G}_{kc}\right] \qquad (14, 4.2)$$

donde

$$\hat{G}_j = \sum_{k\in s} d_k\, \Delta(t - \tilde{\alpha} - \tilde{\beta} x_j - \tilde{\epsilon}_k) / \sum_{k\in s} d_k \qquad (15, 4.2)$$

$$\hat{G}_{kc} = \sum_{i\in s} \frac{\pi_k}{\pi_{ki}} \Delta(t - \tilde{\alpha} - \tilde{\beta} x_k - \tilde{\epsilon}_i) / \sum_{i\in s} \frac{\pi_k}{\pi_{ki}} \qquad (16, 4.2)$$

donde

$$\tilde{\beta} = \frac{\sum_{k\in s} d_k (x_k - \tilde{x}\;)\,(y_k - \tilde{y}\;)}{\sum_{k\in s} d_k (x_k - \tilde{x})^2} \qquad (17, 4.2)$$

$\tilde{\alpha} = \tilde{y} - \tilde{\beta}\tilde{x}$, $\tilde{\epsilon}_k = y_k - \tilde{\alpha} - \tilde{\beta} x_k$; y

$$\tilde{y} = \frac{\sum_{k\in s} d_k y_k}{\sum_{k\in s} d_k} \;\; ; \;\; \tilde{x} = \frac{\sum_{k\in s} d_k x_k}{\sum_{k\in s} d_k}$$

El estimador $\hat{F}_{RKM}(t)$, fue motivado como un estimador diferencia, asumiendo que:

$$G_j = \frac{1}{N} \sum_{i=1}^{N} \Delta(t - \alpha - \beta x_j - \epsilon_i) \qquad (18, 4.2)$$

son conocidos y entonces son reemplazados por sus estimaciones.

- ✓ Estimadores de calibración:

El método de calibración (Deville y Särndal, 1992) fue originalmente empleado en la estimación de la media y más recientemente (Rueda y otros 2007; Martinez y otros 2010) ha sido aplicado a la estimación de la función de distribución. Este método, aplicado a la estimación de la función de distribución, sustituye los pesos básicos d_i por unos nuevos pesos ω_i, que minimizan la distancia

$$\phi_s = \frac{1}{2} \sum_{i\in s} \frac{(\omega_i - d_i)^2}{d_i q_i} \qquad (19, 4.2)$$

sujeto a las siguientes condiciones de calibración:

$$\frac{1}{N} \cdot \sum_{i\in U} \omega_i \Delta(t_1 - g_i) = F_g(t_1) \qquad (20, 4.2)$$

$$\frac{1}{N} \cdot \sum_{i\in U} \omega_i \Delta(t_1 - g_i) = F_g(t_2) \qquad (21, 4.2)$$

$$\vdots$$

$$\frac{1}{N} \cdot \sum_{i\in U} \omega_i \Delta(t_1 - g_i) = F_g(t_p) \qquad (22, 4.2)$$

donde los valores q_i son constantes positivas arbitrariamente elegidas y donde :

$t_g = (t_{g1}, t_{g2, \ldots,} t_{gp})$ es un vector de puntos.

Como hemos dicho anteriormente, este estimador de calibración consiste en sustituir en el estimador de Horvitz-Thompson los pesos básicos d_i por unos pesos nuevos ω_i que cumplen dos condiciones:

1. Están próximos a los pesos básicos d_i y conservar la insesgadez.
2. Proporcionan estimaciones perfectas de la función de distribución de la variable auxiliar g en algunos puntos.

Como el objetivo es la línea de pobreza $r_{0.95,0.5} = \frac{Q_y(0.95)}{Q_y(0.5)}$ parece lógico escoger:

$$\frac{1}{N}\sum_{i\in s}\omega_i\Delta(tg_{50}-g_i) = F_g(tg_{50}) \qquad (23, 4.2)$$

$$\frac{1}{N}\sum_{i\in s}\omega_i\Delta(tg_{95}-g_i) = F_g(tg_{95}) \qquad (24, 4.2)$$

El estimador así obtenido tiene la expresión

$$\hat{F}_{YC1}(t) = \hat{F}_{YHT}(t) + \left[F_g(tg) - \hat{F}_{GHT}(tg)\right]' \cdot B \qquad (25, 4.2)$$

donde tg'=(tg_{50}, tg_{59}) y donde

$$\left[F_g(tg) - \hat{F}_{GHT}(tg)\right]' = \left[\left(F_g(tg_{50}) - \hat{F}_{GHT}(tg_{50})\right), \left(F_g(tg_{95}) - \hat{F}_{GHT}(tg_{95})\right)\right]$$

$$B = \left(\sum_{i\in s} d_i q_i \Delta(tg \quad - g_i)\cdot\Delta(tg \quad - g_i)'\right)^{-1}\cdot\left(\sum_{i\in s} d_i q_i \Delta(tg \quad - g_i)\cdot\Delta(t - y_i)\right)$$

supuesto que la matriz

$$\left(\sum_{i\in s} d_i q_i \Delta(tg \quad - g_i)\cdot\Delta(tg \quad - g_i)'\right)$$

es no singular

Otra opción (para representar los ingresos más bajos) es:

$$\frac{1}{N}\sum_{i\in s}\omega_i\Delta(tg_{25}-g_i) = F_g(tg_{25}) \qquad (26, 4.2)$$

$$\frac{1}{N}\sum_{i\in s}\omega_i\Delta(tg_{50}-g_i) = F_g(tg_{50}) \qquad (27, 4.2)$$

$$\frac{1}{N}\sum_{i\in s}\omega_i\Delta(tg_{95}-gi \quad) = F_g(tg_{95}) \qquad (28, 4.2)$$

El estimador obtenido con estas condiciones viene dado por:

$$\hat{F}_{YC2}(t) = \hat{F}_{YHT}(t) + \left[F_g(tg) - \hat{F}_{GHT}(tg)\right]' \cdot B \qquad (29, 4.2)$$

pero en esta ocasión debemos sustituir $tg = (tg_{25}, tg_{50}, tg_{95})$ en las expresiones de $\left[F_g(tg) - \hat{F}_{GHT}(tg)\right]'$ y en B

Finalmente, otra opción que podemos considerar es incluir en las condiciones de calibración de los dos últimos estimadores la condición

$$\frac{1}{N}\sum_{i\in s}\omega_i \Delta(tg_M - g_i) = F_g(tg_M)\text{=1}$$

donde tg_M es el valor máximo de la variable g en la población U. La razón de incluir esta condición se debe a que los estimadores obtenidos con ella son auténticas funciones de distribución, lo que no ocurría con los estimadores $\hat{F}_{YC1}(t)$ y $\hat{F}_{YC2}(t)$.

De este modo obtenemos dos estimadores nuevos cuyas expresiones son:

$$\hat{F}_{YC3}(t) = \hat{F}_{YHT}(t) + \left[F_g(tg) - \hat{F}_{GHT}(tg)\right]' \cdot B \qquad (30, 4.2)$$

donde en esta ocasión el vector $tg = (tg_{25}, tg_{50}, tg_M)$, y el estimador

$$\hat{F}_{YC3}(t) = \hat{F}_{YHT}(t) + \left[F_g(tg) - \hat{F}_{GHT}(tg)\right]' \cdot B \qquad (31, 4.2)$$

donde $tg = (tg_{25}, tg_{50}, tg_{50}, tg_M)$

4.3 Estimadores para Cuantiles y Líneas de Pobreza.

Cuando tenemos información auxiliar disponible, no es claro el procedimiento que se debe seguir para incorporar dicha información auxiliar a la hora de definir nuevos estimadores de los cuantiles. La incorporación de la información auxiliar puede hacerse de dos formas distintas:

1. Construir estimadores de los cuantiles extendiendo las técnicas de razón, diferencia y regresión, en un sentido similar a las técnicas desarrolladas para la media poblacional.
2. Construir estimadores de la función de distribución y después haciendo uso de su monotonía, tomar la inversa en el correspondiente punto para obtener una estimación del cuantil deseado.

De las dos alternativas existentes, la que nos va a interesar es la segunda, ya que la idea es emplear los estimadores para la función de distribución revisados en los apartados anteriores.

El correspondiente cuantil de orden β de la variable y, en la población finita es:

$$Q_y(t) = F_y^{-1}(\beta) = \inf\{t: F_y(t) \geq \beta\} \quad (1, 4.3)$$

y como medida de pobreza emplearemos el ratio 95/50, dado por:

$$r_{0.95,0.5} = \frac{Q_y(0.95)}{Q_y(0.5)} \quad (2, 4.3)$$

El procedimiento general que vamos a emplear es el siguiente:

Primero, obtenemos un estimador de la función de distribución, $\hat{F}_y(t)$, que debe ser monótono no decreciente y entonces se estima el cuantil tomando la inversa, esto es:

$$\hat{F}_y(t) \longrightarrow \hat{Q}_y(\beta)$$

$$\hat{Q}_y(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (3, 4.3)$$

donde $\hat{F}_y^{-1}$ es la función inversa de $\hat{F}_y$.

$$\hat{Q}_{YHT}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (4, 4.3)$$

$$\hat{Q}_R(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (5, 4.3)$$

$$\hat{Q}_d(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (6, 4.3)$$

$$\hat{Q}_{CD}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (7, 4.3)$$

$$\hat{Q}_{RKM}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (8, 4.3)$$

$$\hat{Q}_{YC1}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (9, 4.3)$$

$$\hat{Q}_{YC2}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (10, 4.3)$$

$$\hat{Q}_{YC3}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (11, 4.3)$$

$$\hat{Q}_{YC4}(\beta) = \hat{F}_y^{-1}(\beta) = \inf\{t: \hat{F}_y(t) \geq \beta\} \quad (12, 4.3)$$

Una vez, que disponemos de estimadores para la estimación de cuantiles, podemos definir los estimadores para la línea de pobreza $r_{0.95,0.5}$, que vienen dados por:

$$\hat{r}^{YHT}_{0.95,0.5} = \frac{\hat{Q}_{YHT}(0.95)}{\hat{Q}_{YHT}(0.5)} = \frac{inf\{t:\hat{F}_{YHT}(t)\geq 0.95\}}{inf\{t:\hat{F}_{YHT}(t)\geq 0.5\}} \qquad (13, 4.3)$$

$$\hat{r}^{R}_{0.95,0.5} = \frac{\hat{Q}_{R}(0.95)}{\hat{Q}_{R}(0.5)} = \frac{inf\{t:\hat{F}_{R}(t)\geq 0.95\}}{inf\{t:\hat{F}_{R}(t)\geq 0.5\}} \qquad (14, 4.3)$$

$$\hat{r}^{d}_{0.95,0.5} = \frac{\hat{Q}_{d}(0.95)}{\hat{Q}_{d}(0.5)} = \frac{inf\{t:\hat{F}_{d}(t)\geq 0.95\}}{inf\{t:\hat{F}_{d}(t)\geq 0.5\}} \qquad (15, 4.3)$$

$$\hat{r}^{CD}_{0.95,0.5} = \frac{\hat{Q}_{CD}(0.95)}{\hat{Q}_{CD}(0.5)} = \frac{inf\{t:\hat{F}_{CD}(t)\geq 0.95\}}{inf\{t:\hat{F}_{CD}(t)\geq 0.5\}} \qquad (16, 4.3)$$

$$\hat{r}^{RKM}_{0.95,0.5} = \frac{\hat{Q}_{RKM}(0.95)}{\hat{Q}_{RKM}(0.5)} = \frac{inf\{t:\hat{F}_{RKM}(t)\geq 0.95\}}{inf\{t:\hat{F}_{RKM}(t)\geq 0.5\}} \qquad (17, 4.3)$$

$$\hat{r}^{YC1}_{0.95,0.5} = \frac{\hat{Q}_{YC1}(0.95)}{\hat{Q}_{YC1}(0.5)} = \frac{inf\{t:\hat{F}_{YC1}(t)\geq 0.95\}}{inf\{t:\hat{F}_{YC1}(t)\geq 0.5\}} \qquad (18, 4.3)$$

$$\hat{r}^{YC2}_{0.95,0.5} = \frac{\hat{Q}_{YC2}(0.95)}{\hat{Q}_{YC2}(0.5)} = \frac{inf\{t:\hat{F}_{YC2}(t)\geq 0.95\}}{inf\{t:\hat{F}_{YC2}(t)\geq 0.5\}} \qquad (19, 4.3)$$

$$\hat{r}^{YC3}_{0.95,0.5} = \frac{\hat{Q}_{YC3}(0.95)}{\hat{Q}_{YC3}(0.5)} = \frac{inf\{t:\hat{F}_{YC3}(t)\geq 0.95\}}{inf\{t:\hat{F}_{YC3}(t)\geq 0.5\}} \qquad (20, 4.3)$$

$$\hat{r}^{YC4}_{0.95,0.5} = \frac{\hat{Q}_{YC3}(0.95)}{\hat{Q}_{YC3}(0.5)} = \frac{inf\{t:\hat{F}_{YC3}(t)\geq 0.95\}}{inf\{t:\hat{F}_{YC3}(t)\geq 0.5\}} \qquad (21, 4.3)$$

4.4 Estudio de Simulación.

En este apartado vamos a ilustrar el comportamiento de los estimadores de medidas de pobreza analizados en el apartado anterior mediante estudios de simulación, que han sido programados con el programa estadístico Rgui.

Concretamente, el estudio se ha llevado a cabo con una población de datos reales correspondientes a N=10000 familias españolas, que han sido extraídos de la Encuesta de Presupuestos Familiares realizada por el Instituto Nacional de Estadística (INE) correspondiente al año 2012 y donde hemos considerado como variable de estudio *y*="Ingresos mensuales familiares" y donde solamente hemos considerado una variable auxiliar *x*="Gastos totales anuales familiares". Para llevar a cabo el estudio hemos considerado los tres diseños muestrales descritos en el apartado 4.1, esto es:

- Muestreo Aleatorio Simple (MAS)
- Muestreo Sistemático
- Muestreo Estratificado

donde en el muestreo estratificado hemos empleado como estratos las comunidades autónomas a la que pertenece cada familia.

En cada uno de los diseños muestrales, se seleccionaron 10000 muestras para cuatro tamaños muestrales distintos, concretamente los tamaños considerados han sido n=200, 300, 400 y 500. De este modo, con cada estimador realizamos 10000 estimaciones de la línea de pobreza $r_{0.95,0.5}$ con cada uno de los tamaños muestrales.

Para evaluar el comportamiento de los estimadores propuestos después de haber realizado las simulaciones hemos considerado el sesgo relativo (RB)

$$RB\left(\hat{R}_{0.95,0.5}\right) = \frac{1}{R_{0.95,0.5}}\left[\frac{1}{10000}\sum_{r=1}^{10000}\left(\hat{R}_{0.95,0.5}{}^{(r)} - R_{0.95,0.5}\right)\right] \qquad (1, 4.4)$$

y la eficiencia relativa (RE)

$$RE\left(\hat{R}_{0.95,0.5}\right) = \frac{ECM\left(\hat{R}_{0.95,0.5}\right)}{ECM\left(\hat{r}_{0.95,0.5}^{YHT}\right)} \qquad (2, 4.4)$$

siendo ECM el error cuadrático medio de un estimador definido como:

$$ECM\left(\hat{r}_{0.95,0.5}\right) = \frac{1}{10000}\sum_{r=1}^{10000}\left(\hat{r}_{0.95,0.5}{}^{(r)} - r_{0.95,0.5}\right)^2 \qquad (3, 4.4)$$

donde $\hat{r}_{0.95,0.5}$ es un estimador cualquiera de la línea de pobreza y donde $\hat{r}_{0.95,0.5}{}^{(r)}$ es la estimación r-ésima realizada con el estimador $\hat{r}_{0.95,0.5}$. La eficiencia relativa RE es la eficiencia relativa de cada estimador con respecto al estimador de Horvitz-Thompson. Un valor de RE>100 significa que el estimador es infeciente con respecto al estimador de Horvitz-Thompson. En la tabla 1 podemos examinar los resultados obtenidos con muestreo aleatorio simple para todos los tamaños de muestra considerados

Tabla (1, 4.4)

Tamaño de muestra n=200									
Estimador	YHT	R	D	CD	RKM	YC1	YC2	YC3	YC4
RB	-0.007	-0.011	0.0014	0.0005	0.005	0.016	0.016	0.004	0.003
RE	1.00	1.12	1.13	0.78	0.79	0.96	0.92	0.78	0.77
Tamaño de muestra n=300									
RB	-0.013	-0.016	-0.003	-0.007	-0.004	-0.005	-0.005	0.001	-0.001
RE	1.00	1.06	1.04	0.65	0.68	0.75	0.73	0.63	0.62
Tamaño de muestra n=400									
RB	-0.006	-0.009	0.004	0.003	-0.004	-0.003	-0.003	0.0009	0.0008
RE	1.00	0.98	0.97	0.60	0.59	0.64	0.62	0.58	0.52
Tamaño de muestra n=500									
RB	-0.005	-0.01	0.003	0.0009	0.0009	0.004	0.003	0.0007	0.0006
RE	1.00	0.92	0.87	0.53	0.54	0.58	0.56	0.51	0.49

En los resultados obtenidos con muestreo aleatorio simple, podemos destacar que el sesgo relativo de los estimadores considerados es prácticamente despreciable, y en general son los estimadores de razón, diferencia y Horvitz-Thompson los que producen unos sesgos mayores. Los estimadores calibrados YC3 e YC4 son los que presentan un sesgo, en general, más pequeño, seguido de los estimadores de Chambers y Dunstan y Rao-Kovar-Mantel. En cuanto a la eficiencia relativa, podemos observar que, salvo raras ocasiones, los estimadores indirectos considerados ofrecen una eficiencia mayor que el estimador de Horvitz-Thompson, y este hecho es más apreciable cuanto mayor es el tamaño muestral considerado. Los estimadores con una mayor eficiencia son los estimadores calibrados YC3 e YC4, que presentan una ganancia en eficiencia con respecto al estimador de Horvitz-Thompson del 22% en el peor de los casos (tamaño n=200) y una ganacia del 51% en el mejor de ellos (tamaño n=500). De este modo, estos estimadores hacen un uso más eficaz de la información auxiliar que el resto de estimadores. Los estimadores de Chambers y Dunstan y Rao-Kovar-Mantel presentan un buen comportamiento en eficiencia, con resultados similares a los estimadores calibrados YC3 e YC4, aunque en general siempre con una menor ganancia en eficiencia que estos últimos. También podemos destacar que los estimadores que menos ganancia en eficiencia presentan son los estimadores de razón y el de diferencia, que suelen presentar una ganancia en eficiencia en torno al 10% y en ocasiones son más ineficientes que el estimador de Horvitz-Thompson, lo que nos lleva a la conclusión de que estos estimadores incorporan de una manera menos eficaz que el resto de estimadores considerados.

La tabla 2 muestra los resultados para muestreo sistemático. En ella podemos observar a grandes rasgos el mismo comportamiento que el muestreo aleatorio simple, esto es, la mayoría de estimadores presenta un sesgo relativo prácticamente despreciable y la mayoría de ellos presenta una eficiencia mayor que el estimador de Horvitz-Thompson, si bien la ganancia en eficiencia, con este tipo de muestreo es, en general, menor que la obtenida con muestreo aletorio simple. Los estimadores más eficientes vuelven a ser los estimadores calibrados YC3 e YC4,con una ganancia en eficiencie que el mejor de los casos alcanza un 20%. Los estimadores menos eficientes vuelven a ser los estimadores de razón y diferencia, y en la mayoría de casos suelen ser menos eficientes que el estimador de Horvitz-Thompson.

Tabla (2, 4.4)

Tamaño de muestra n=200									
Estimador	YHT	R	D	CD	RKM	YC1	YC2	YC3	YC4
RB	0.008	0.013	0.012	0.009	0.010	0.011	0.009	0.007	0.007
RE	1.00	1.22	1.18	0.96	0.92	0.99	0.98	0.92	0.90
Tamaño de muestra n=300									
RB	0.006	0.011	0.011	-0.006	-0.007	-0.008	-0.006	-0.004	0.003
RE	1.00	1.15	1.11	0.92	0.90	0.96	0.95	0.88	0.87
Tamaño de muestra n=400									
RB	0.003	-0.005	-0.007	0.001	-0.001	-0.002	-0.003	0.0009	0.0008
RE	1.00	1.07	1.03	0.88	0.87	0.91	0.90	0.86	0.85

Tamaño de muestra n=500									
RB	0.003	-0.005	-0.007	0.001	-0.001	-0.002	-0.003	0.0009	0.0008
RE	1.00	0.99	0.98	0.86	0.86	0.90	0.89	0.81	0.80

En los resultados correspondientes a muestreo estratificado (Tabla 3) , podemos observar un comportamiento prácticamente similar al obtenido en muestreo aleatorio simple, si bien en este tipo de muestreo la ganancia en eficiencia es mayor que la obtenida con el muestreo aleatorio simple. Así, los sesgos siguen siendo despreciables para todos los estimadores considerados y podemos ver que los estimadores que ofrecen un resultados mejor en cuanto a eficiencia relativa son el estimador YC4 y el estimador de Chambers y Dunstan, con una ganancia en eficiencia que el mejor de los casos es del 60%. Los estimadores de Rao-Kovar-Mantel y el estimador calibrado YC3 también tienen un comportamiento similar entre ellos con una ganancia en eficiencia del 52%, siendo nuevamente los estimadores de razón y diferencia los que ofrecen una peor eficiencia relativa, si bien con el muestreo estratificado ambos estimadores siempre son más eficientes que el estimador de Horvitz-Thompson, con lo que con el muestreo estratificado todos los estimadores que incorporan información auxiliar producen mejores resultados que el estimador de Horvitz-Thompson, que recordemos no hace uso de esta información.

Tabla (3, 4.4)

Tamaño de muestra n=200									
Estimador	YHT	R	D	CD	RKM	YC1	YC2	YC3	YC4
RB	0.006	0.010	0.002	-0.001	-0.001	0.002	0.003	0.002	0.003
RE	1.00	0.96	0.97	0.73	0.74	0.83	0.82	0.75	0.71
Tamaño de muestra n=300									
RB	0.003	-0.008	0.001	0.0006	0.0008	0.001	0.0009	0.0006	0.0005
RE	1.00	0.90	0.91	0.65	0.68	0.73	0.72	0.69	0.64

Tamaño de muestra n=400									
RB	0.001	-0.004	-0.002	0.0003	0.0005	0.0009	0.0007	0.0005	0.0003
RE	1.00	0.86	0.82	0.53	0.56	0.67	0.68	0.57	0.54
Tamaño de muestra n=500									
RB	0.0009	0.002	0.002	0.0002	0.0004	0.0007	0.0004	0.0001	0.0001
RE	1.00	0.81	0.78	0.41	0.49	0.53	0.55	0.48	0.40

5. Conclusiones

En el presente trabajo se ha realizado una revisión de los diferentes enfoques de la pobreza y de las diferentes medidas empleadas para el análisis de la misma, centrando su estudio en aquellas medidas basadas en ratios de percentiles y más concretamente en el problema de la estimación de tales medidas a través de estudios muestrales.

De este modo, el objetivo principal que persigue el trabajo es abordar el problema de la estimación de las medidas de pobreza de una población finita basadas en ratios de percentiles cuando sólo disponemos de la información proporcionada por una muestra y más concretamente cómo incorporar la información auxiliar disponible en el proceso de estimación, empleando la opción de incorporar dicha información en la estimación de la función de distribución.

D este modo, en el trabajo se ha realizado una revisión de los principales estimadores indirectos de la función de distribución, como son los estimadores de Chambers y Dunstan, el estimador de Rao, Kovar y Mantel y técnicas más novedosas como el método de calibración para la incorporación de información auxiliar. Para todos ellos, se han analizado sus propiedades teóricas y además se ha realizado un ejemplo práctico de cómo emplear estos estimadores con datos reales mediante un estudio de simulación con datos obtenidos de la Encuesta de Presupuestos Familiares realizada por el Instituto Nacional de Estadística (INE) correspondiente al año 2012, donde podemos conprobar la eficiencia de los estimadores indirectos, es decir, la incorporación de la información auxiliar produce mejores estimaciones de las medidas de pobreza consideradas.

6. Bibliografía

Chambers, R.L., & Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika 73*, 597–604.

Deville, J.C., & Särndal, C.E. (1992). Calibration estimators in survey sampling. *J. Am. Stat. Assoc. 87*, 376–382.

Dickens, R., & Manning, A. (2004). Has the National Minimum Wage Reduced UK Wage Inequality?. *Journal of the Royal Statistical Society. Series A, 167,* 613-626.

European Commission. (1998). *Social protection in Europe. European Commission*. Brussels.

Fernández, F. R., & Mayor, J. A. (1995). *Muestreo en poblaciones finitas: curso básico.* Barcelona, Ediciones Universitarias de Barcelona (EUB).

Instituto Nacional de Estadística (INE). *La pobreza y su medición.*

Kahn, L. (1998) Collective bargaining and the interindustry wage structure: international evidence. *Economica 65*, 507–534.

Martínez, S., Rueda, M., Arcos, A., & Martínez, H. (2010). Optimum calibration points estimating distribution functions. *J. Comput. Appl. Math. 233*, 2265–2277.

Martínez, S., & Martínez, H. (2013). Aplicación de técnicas de calibración en la estimación de líneas de pobreza. *Estadística Española, 55*, 323-335

Rao, J.N.K., Kovar, J.G., & Mantel, H.J. (1990). On estimating distribution function and quantiles from survey data using auxiliary information. *Biometrika 77*, 365–375.

Rueda, M., Martínez, S., Martínez, H., & Arcos, A. (2007). Estimation of the distribution function with calibration methods. *J. Stat. Plan. Inf. 137*, 435–448.

Rueda, M., & Muñoz, J.F. (2011). Estimation of poverty measures with auxiliary information in sample surveys. *Quality & Quantity, 45,* 687-700. doi: 10.1007/s11135-009-9279-y.

Särndal, C.E. (2007). The calibration approach in survey theory and practice. *Survey Methodology, 33*, 99-119.

Singh, S. (2003). *Advanced sampling theory with applications: How Michael ”selected”*. The Netherlands, Amy. Kluwer Academic Publisher.