Elda Gallese
Nora Mabel Lac Prugent

Jefe de colaboradores: Norberto Martín
Colaboradores: Leonardo Jorge Balestro, Julián Ignacio Crucella, María Eugenia Fernández de Luco, Rodrigo Daniel López, María Valeria Paccapelo y Nicolás Omar Renzi.
Instituto de Investigaciones Económicas, Escuela de Economía
Consejo de Investigaciones de la Universidad Nacional de Rosario

INGRESOS. OBTENCIÓN DE DATOS CONFIABLES Y MODELADO EN EL CONTEXTO DEL TERCER MILENIO[1]

1 INTRODUCCIÓN

En octubre del 2003 el presidente de Argentina presentaba el documento "Objetivos de Desarrollo del Milenio. Argentina. La oportunidad para su reencuentro" con estas palabras:

> …En el año 2000 todos los países del mundo firmamos un convenio sobre los Objetivos de Desarrollo del Milenio que deberíamos alcanzar en el 2015. Ellos reflejan la factibilidad a partir de los recursos tecnológicos y de conocimiento que la humanidad dispone, de reducir la pobreza extrema y el hambre, así como mejorar la cobertura, calidad y equidad en educación y salud, realizándolo con políticas de desarrollo sostenible y promoviendo valores de equidad y solidaridad de género, generacionales y territoriales…

El primer objetivo a alcanzar, "erradicar la pobreza extrema y el hambre", se debería lograr a través del cumplimiento de dos metas:

Meta 1: erradicar la indigencia y el hambre.

Meta 2: Reducir la pobreza a menos del 20%.

También dice el presidente en la presentación:

> …Para lograrlo todos debemos aportar. El Gobierno, promoviendo las mejores políticas a nivel nacional, provincial y municipal. La sociedad participando y controlando. Los empresarios, invirtiendo, generando empleo y cumpliendo con sus obligaciones fiscales. Los trabajadores sumando su esfuerzo y su creatividad al éxito de las organizaciones privadas y públicas en las que se desempeñan. Los niños y los jóvenes estudiando y preparándose…

Ya que la tarea de la ciencia es la de interpretar la realidad y no suplantarla, lograr la estimación confiable de la Distribución de los Ingresos, ayudaría a realizar el seguimiento del primer objetivo, satisfaciendo el pedido del presidente hacia la sociedad, de participar y controlar.

En la producción de datos, uno de los errores más perturbadores ajenos al diseño de la muestra es el debido a errores de respuestas, ya sea por no respuestas o respuestas falseadas. Tal ese el caso de la base de datos que suministra la Encuesta Permanente a Hogares (EPH), cuya tasa de no respuestas en la pregunta referida al Ingreso alcanza niveles alarmantes, superando en las últimas captaciones el treinta por ciento.

En este trabajo se presentarán algunos modelos para intentar lograr una estimación más confiable de la Distribución de los Ingresos.

---

[1] Trabajo realizado en el marco del Proyecto de Investigación "¿Se están alcanzando los objetivos de desarrollo del milenio?". Programa 202 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Rosario. Dirección: Nora Lac Prugent. [1ECO62]

Son métodos de estimación modelados estadísticamente. Se pondrá especial atención a los modelos de obtención de datos que disminuyan la probabilidad de respuestas falseadas.

La estructura de este trabajo tendrá el siguiente desarrollo:

En la introducción quedó planteado el problema (poca confiabilidad en los datos sobre ingresos). Se detallarán los objetivos a continuación. En el tercer y cuarto apartado se contextualizará el tema a tratar. En el tres, el ingreso disponible se mostrará dentro de un esquema simplificado de una economía nacional. En el cuarto dentro de la dinámica de la investigación aplicada; “la obtención de datos confiables” y “análisis de los mismos” recorrerán el Ciclo investigativo y los tipos de pensamientos estadísticos involucrados en esta dinámica.

A través del tiempo se idearon diversos modelos aleatorizados de interrogación, algunos de los cuales serán tratados sucintamente en el apartado cinco. En la sección seis se desarrollará la técnica para el relevamiento de datos referidos a modelos tricotómicos y en la séptima una adaptación a la obtención del histograma del Ingreso de las personas. Se presentarán algunos modelos aptos para el análisis de estos datos y los métodos de ajuste de funciones en la sección ocho. Finalmente se esbozará algunas conclusiones y reflexiones en el último apartado.

2 OBJETIVOS

Propósito:

El propósito didáctico de este trabajo, como el de otros de la misma autoría, es el de encaminarse hacia el logro de algunas metas.

1-La de fomentar en los jóvenes investigadores, “las ganas” de intentar establecer un nexo entre la producción de datos de calidad, la investigación científica y la enseñanza.

2-La de construir un instrumento pedagógico, con respecto a la dinámica de la investigación, destinado principalmente a capacitar a los alumnos.

El objetivo central *de este trabajo es el de acercar posibles soluciones al complejo problema de la Distribución de los Ingresos de las personas.*

Se espera alcanzar este objetivo mediante:

a-La presentación y desarrollo de algunos métodos probabilísticos para la captación de los datos referidos al ingreso de las personas. Este será el de mayor interés.

Estos métodos (producción de datos) son tales que producen datos que transportan información resguardando la privacidad.

En general, en estudios que tienen por objetivo el análisis de los ingresos de una población, se pueden esperar negativas o respuestas evasivas por parte del entrevistado. La falta de respuesta produce, por un lado, una reducción del tamaño muestral que disminuye la precisión y por otro, un sesgo independiente de dicho tamaño.

Se han propuesto distintos métodos para el tratamiento de la falta de respuesta. Para minimizar el efecto de la no respuesta de ingresos, la EPH asigna a los no respondentes el comportamiento de los respondentes por estrato de la muestra. ¿Nada puede hacerse para paliar el efecto de las respuestas falseadas?

Para intentar dar respuesta a este interrogante, en 1976 Gallese realiza una adaptación de un modelo aleatorizado cualitativo al problema de la estimación de características numéricas continuas, ejemplificándolo con el problema de la “distribución de la renta de las personas”.

Actualmente, inmersos en el tercer milenio, métodos similares se utilizan para preservar la privacidad de la información. A las múltiples macro matrices de datos que circulan por el cyber espacio, se los “perturba” con bombardeos probabilísticos de suerte que no sea posible reconstruir la información brindada individualmente. Sin embargo, esta perturbación es tal que la base de datos preserva su potencial estadístico (Dinur y Nissim 2003).

b-La presentación de algunos modelos que describen la Distribución de los Ingresos entre las personas.

Los modelos (herramientas del análisis) son aquellos con posibilidades potenciales de interpretar el comportamiento de la Distribución de los Ingresos, entre las personas o entre las familias, a partir de datos agrupados. Primera y segunda ley de Pareto, la ecuación de Amoroso y la función de Dagum.

c-El desarrollo de dos métodos de ajuste de funciones (herramientas del análisis), que permitirían aventurar el comportamiento poblacional a partir de la ventana de observación, la muestra.

De los dos métodos desarrollados para el ajuste de funciones, uno el "método de los momentos", ajusta la función de densidad aplicada a la ecuación de Amoroso y el otro, el "método del punto fijo" ajusta la función de distribución acumulada de Dagum.

3 CONTEXTO. DISEÑO ANALÓGICO DE UNA ECONOMÍA NACIONAL SIMPLIFICADO

Ubicación del ingreso disponible en la economía nacional.

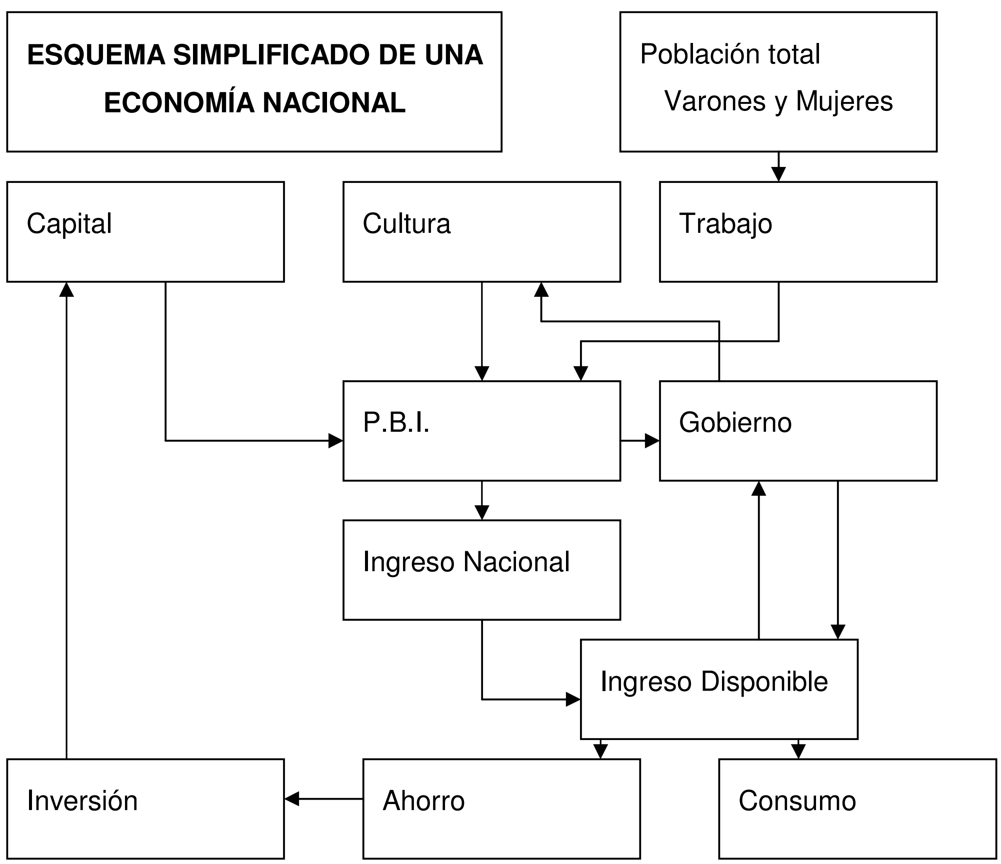


Figura 1. Esquema simplificado de una economía nacional

La figura 1 muestra un ideograma que actuará como un mapa de los distintos puntos que se podrían recorrer. En este trabajo el foco estará puesto en la “distribución de los ingresos disponibles de las familias”.

Sea Q el símbolo del capital productivo, es decir: los campos con sus enceres, máquinas agrícolas, etc., las plantas industriales con sus terrenos, edificios, maquinarias, materias a procesar, etc., los comercios y demás instituciones, es decir todas las fuentes de producción de bienes y servicios que se medirán a los precios de mercado. Si imaginamos que el valor de Q lo asimilamos a un líquido, en el comienzo de un período –generalmente un año- se lo podrá representar con el contenido de una vasija.

Es evidente que en el proceso de la producción el nivel –el capital- irá descendiendo, pero por otra parte se irá compensando lo consumido –y debe ser positivamente para que la economía esté en salud con las inversiones.

Esto indica que en el recipiente habrá dos grifos: uno en el que el líquido sale y otro por el que se incorpora a la vasija.

Q es una de las variables de la macroeconomía más difícil de medir. Por ejemplo, la maquinaria de una industria tiene un valor histórico cuando se las adquirió y que luego se ha visto disminuido por su depreciación. Pero el valor de mercado es muy difícil de medir. En Gran Bretaña, por ejemplo, su personal técnico, puede brindar una buena información.

Se tratará de dar una primera visión del ideograma general de una macroeconomía, sin olvidar que los problemas económicos son demasiado complejos como para que pueda explicarse por este esquema. Simplemente sirve para dar una idea de la relación entre los distintos sectores que intervienen en la economía nacional.

El Producto Bruto Interno (PIB)[2] es el valor de los bienes y servicios que se han ido produciendo a lo largo de un año, medidos a precio de mercado. Es un concepto dinámico. De velocidad. Es la diferencia de la Riqueza de un año con respecto al anterior. Crecimiento de la riqueza en la unidad de tiempo. En los precios de mercado están incluidos los impuestos indirectos. Esto implica que del PIB salga dinero para el gobierno (G.)

L simboliza la población económicamente activa, la que participa de la producción tanto de bienes como de servicios. Se mide en horas de trabajo. Es una letra que viene del inglés (labor). L es parte de N, la población total, como N crece y entran con L nuevos elementos al mercado de trabajo ya se ve que hay que controlar el equilibrio de un crecimiento armonizado de L y Q.

K significa la cultura, los tres tipos de enseñanza que al capacitar a los elementos de L elevan la calidad y el volumen del producto P.

P depende entonces de Q, de L, y de K

$$P = F (Q, L, K)$$

El Ingreso Nacional distribuido entre las personas constituye el Ingreso disponible. De la vasija de ingreso disponible sale dinero para el gobierno (G) por medio de los entes recaudadores de impuestos. y también ingresa dinero disponible del gobierno por medio de jubilaciones y pensiones.

---

[2] El profesor Kustner de la Universidad de de Harvard es un experto de estos problemas del producto y la O.N.U. ha tenido en cuenta muchas de sus recomendaciones para dar normas en muchos países. Los primeros cálculos del PIB en la Argentina los hizo Loreto Domínguez alrededor del año 1940. Posteriormente el equipo lo tomó Manuel Balboa y luego Fracchia. En mayo de 1955 se publica un libro del Producto Bruto estimando las cuentas desde el año 1935 hasta el año 1954.

Del gobierno sale dinero destinado a elevar la cultura de la población. Esta elevación de la cultura contribuye a elevar el Producto Bruto mediante la elevación de la calidad y volumen del producto.

Del ingreso disponible (tema a desarrollar en este trabajo) una parte va destinada al consumo (C) (de bienes perecederos y bienes durables) y otra parte va destinada al ahorro (S) que –suponiendo que no hay atesoramiento- por medio de los resortes bancarios constituye la inversión bruta, o sea reposición de los bienes de capital gastados en el proceso de producción más la inversión neta. La inversión neta debe ser positiva para que la economía esté en salud.

Se puede vislumbrar en este esquema que la teoría de la Distribución de los Ingresos comprende dos aspectos diferentes pero interdependientes: uno considera el mecanismo remunerativo de los factores de la producción, el otro los efectos de esa remuneración a través de la distribución de los ingresos entre los distintos grupos sociales. Es este último el aspecto que se aborda en este trabajo.

4 EL PENSAMIENTO ESTADÍSTICO EN INVESTIGACIONES EMPÍRICAS

En la esta sección se presenta un resumen del artículo "Statistical Thinking in Empirical Enquiry"[3]En el mismo se discute el proceso de pensamiento involucrado en la resolución de problemas estadísticos, desde el momento de su formulación hasta el arribo de conclusiones.

Se identifica un sistema de cuatro dimensiones que incluye *un ciclo investigativo, un ciclo interrogativo, tipos de pensamiento y disposiciones*. El elemento central de este pensamiento es la variación y por lo tanto se discute su rol en la concepción estadística de los problemas del mundo real, incluyendo la búsqueda de sus causas.

4.1 Esquematización del Pensamiento Estadístico en investigaciones empíricas

El esquema que se propone en el artículo involucra a todo aquel que participa de la investigación, no sólo a los estadísticos. El mismo, busca organizar algunos elementos del pensamiento estadístico durante la investigación a través de cuatro dimensiones:

Dimensión 1: Ciclo Investigativo

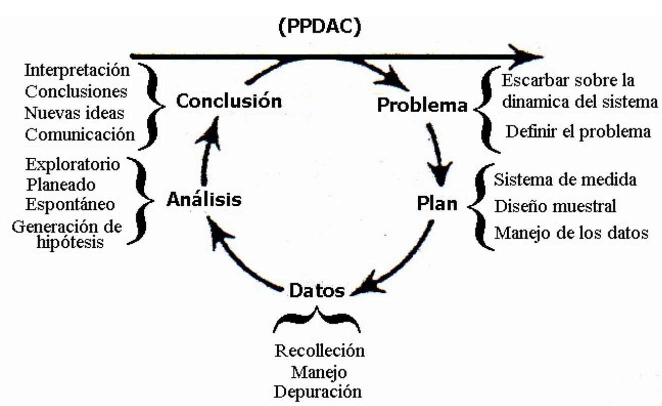


Figura 2: El ciclo investigativo

[3] Traducción y resumen realizados por los alumnos de la Licenciatura en estadística de los años 2004 y 2006, de Wild y Pfannkuch, publicado en la revista científica "International Statistical Review" (1999).

En la Figura 2[4] se intenta esquematizar el modo de actuar y pensar durante el curso de una investigación estadística. Se adapta el modelo PPDAC (Problema, Plan, Datos, Análisis y Conclusiones) de MacKay y Oldford (1994). El ciclo PPDAC concierne a la abstracción y resolución de un problema estadístico a partir de un problema real.

En la primera etapa se profundiza sobre la dinámica del sistema y se define el *problema.* Luego, se *planean* las técnicas a utilizar para su solución, referidas a: el sistema de medida, el diseño muestral y el manejo de datos. Posteriormente, se presenta todo aquello relacionado con la recopilación, administración y depuración de los *datos* necesarios. Se continúa con el *análisis* exploratorio, planeado y espontáneo de los datos y el análisis confirmatorio. Esta instancia también incluye la generación de hipótesis. Finalmente, se interpretan los resultados, se obtienen *conclusiones* y se plantean nuevos interrogantes para continuar en una futura investigación.

Los conocimientos ganados y las necesidades identificadas durante el ciclo PPDAC pueden iniciar futuros ciclos investigativos. Las conclusiones obtenidas a partir de las investigaciones amplían el conocimiento del contexto que puede ser de utilidad para llevar a cabo alguna acción.

Dimensión 2: Tipos de pensamiento

Se distinguen dos tipos de pensamientos, los de carácter general y los fundamentales para el pensamiento estadístico.

Los pensamientos generales aplicados en un contexto estadístico se subdividen en:

- *estratégico*: comprende el planeamiento es decir, la forma en que se aborda el problema y la concientización de las limitaciones prácticas bajo las cuales se trabaja.
- *de búsqueda* de explicaciones: consiste en el estudio de las relaciones causa-efectos y de las posibles fuentes de variación.
- *de modelización*: abarca la construcción de modelos que permiten obtener una representación simplificada del fenómeno en cuestión, con el objetivo de comprender y predecir su comportamiento.
- *aplicación de técnicas*: intenta abordar un problema nuevo a partir del conocimiento de algún otro que haya sido resuelto. Además, comprende el reconocimiento y uso de arquetipos y la utilización de las herramientas disponibles para la resolución del problema.

Entre los tipos fundamentales para el pensamiento estadístico se consideran:

- *reconocimiento de la necesidad de datos.*
- *transnumeración:* comprende el uso de herramientas que aporta la estadística para extraer la información de los datos. Incluye la obtención de mediciones de un sistema real, el cambio en la representación de los datos, análisis exploratorios para detectar patrones de comportamiento, etc. Se trata de un proceso dinámico que facilita el conocimiento.
- *variación*: se refiere al aprendizaje y la toma de decisiones bajo incertidumbre que surge principalmente debido a la variación omnipresente. Es importante la medición y la modelización de ésta con el propósito de predecir, explicar o controlar. En la siguiente sección se desarrolla con mayor detalle.
- *razonamiento con modelos estadísticos*
- *integración entre los conocimientos estadísticos y contextuales*

[4] Extraído de Crucella, J; Fernández de Luco, M. E. y Renzi, N. (2005) *Ajuste de funciones no lineales*. Trabajo ganador del Primer premio “Ingeniero Carbajo”

Dimensión 3: Ciclo interrogativo. Las preguntas en cada nodo del ciclo interrogativo y en cada estadío del pensamiento estadístico son más importantes que las respuestas.

Dimensión 4: Disposiciones. Curiosidad. Tenacidad. Duda. Escepticismo. Siendo lógicos. Imaginación. Búsqueda de significados más precisos.

4.1.1 El ciclo investigativo en esta propuesta

El problema (P): ¿con las altas tasas de no respuesta presente en la EPH, los datos son confiables? ¿Se podrían obtener datos más confiables sobre el ingreso de las personas? ¿Hay modelos para suavizar, minimizar o eliminar el efecto de las respuestas falseadas? ¿Qué modelos y métodos de ajuste de funciones son aptos para datos agrupados?

El Plan (P) Contextualizar el problema. Buscar estrategias para la obtención de datos confiables. Analizar los modelos teóricos para ajuste de la Distribución de los Ingresos y desarrollar métodos de estimación apropiados.

Datos (D): Proponer métodos aleatorizados para recabar datos que trasporten información fidedigna resguardando la privacidad.

Análisis (A): Desarrollar métodos de estimación eficientes para el ajuste de funciones cuando la ventana de observación (la muestra) proporciona el histograma de la Distribución de los Ingresos.

Conclusiones: Seguir investigando sobre la eficiencia de los métodos propuestos.

4.1.2 Los tipos de "pensamiento estadístico" involucrados en este trabajo

Reconocimiento de la necesidad de datos de calidad. Datos que transporten información preservando la intimidad de los individuos.

Presencia de la variación, modelándola desde el propio mecanismo de interrogación

Razonamiento con modelos estadísticos para el análisis.

Integración de los conocimientos estadísticos y contextuales. El problema se ha contextualizado dentro de la teoría económica y los Objetivos del Milenio.

5 RESPUESTAS ALEATORIZADAS: una técnica de investigación para eliminar el vicio estadístico producido por respuestas falseadas

¿Cómo preservar la privacidad manteniendo la información? Warner (1965) elaboró un modelo que permite guardar el anonimato de la pregunta contestada. Esto se logra mediante un refinamiento en el mecanismo de las interrogaciones, construido sobre artificios que introducen algunos sencillos elementos de probabilidad.

Una de las principales preocupaciones de la Estadística es la de tratar de mejorar sustancialmente los métodos de captación, procesamiento y análisis de datos a los efectos de extraer la máxima información. Dicha mejora trae aparejado ahorro de tiempo, mayor rapidez en las operaciones y aumento de la fiabilidad de los resultados, teniendo presente que esto debe alcanzarse al menor costo posible. Pero también debe señalarse que uno de sus objetivos consiste en ir ensanchando los dominios de la información agregando nuevos conocimientos de hechos anteriormente no registrados.

Este ensanchamiento de la capacidad de obtención de información, se torna un tema de suma importancia en el dominio de las disciplinas sociales por cuanto permite penetrar en las características de los diferentes elementos que constituyen una población humana. Los impedimentos que han tenido los métodos tradicionales de captación de información estadística, en los renglones de naturaleza peculiarmente reservada, pueden resolverse superando estas etapas. Esto se logra mediante la incorporación de recursos probabilísticos que aseguren al informante que la información suministrada por él no podrá ser detectada de ningún modo por el entrevistador y por ende, por ninguna persona interviniente en el proceso de la investigación. De igual modo el mecanismo de aleatorización de la pregunta debe ser tal que permita, por medio del análisis estadístico, la estimación posterior de las proporciones o promedios actuantes en la población en estudio.

Interrogaciones Aleatorizadas es una técnica de investigación para eliminar el vicio estadístico producido por respuestas falseadas.

Este método pasó a ser la piedra fundamental sobre la cual se construyeron los diversos modelos aleatorizados, abriendo así un importante camino en un campo de la estadística. Esto es la recolección fidedigna de los datos manteniendo la privacidad de los mismos.

5.1 Errores

Al planear una encuesta el objetivo del investigador debe ser el de minimizar el "error total de la investigación", no solamente controlar el "error aleatorio muestral".

La teoría clásica de los errores considera a los errores sistemáticos y los errores accidentales. Los errores sistemáticos son los que escapan a correcciones analíticas y son controladas mediante revisaciones minuciosas, correcciones físicas, reiteraciones, etc. Los errores accidentales son debido a múltiples y diferentes causas externas, imposibles de controlar y se cree que siguen la ley Normal de Gauss. Esta creencia va cayendo cada vez más en descrédito y se piensa que la ley Normal de Gauss es sólo una idealización de la verdadera ley de los errores accidentales.

Trazando una paralela entre la teoría clásica y la teoría estadística de los errores, a los errores accidentales le correspondería el "error aleatorio muestral", probabilísticamente controlable por medio del armamento que suministra la Teoría de las Muestras. En la práctica, al error aleatorio muestral se lo puede pensar como la diferencia entre dos resultados: uno obtenido por investigación muestral y otro por un requerimiento exhaustivo de la población, utilizando en ambos casos los mismos métodos de recolección, clasificación, tabulación, etc., iguales definiciones y exactamente los mismos recaudos en las interrogaciones, entrenamiento de los entrevistadores, etc.

Los errores sistemáticos pueden ser equiparados a los "errores ajenos al diseño de la muestra". Estos pueden echar por tierra el nivel de precisión esperando aún cuando el diseño de la muestra haya sido eficientemente elaborado.

5.1.1 Errores de respuestas

Los errores de respuesta pueden deberse a diversas causas, pero es obvio que cuando se realizan investigaciones en poblaciones humanas acerca de temas íntimos, delicados, la principal fuente de error de respuesta es debida a la entendible resistencia que todo individuo tiene para revelar a extraños asuntos personales íntimos, tales como: comportamiento delictivo (drogadicción, abortos, etc.), deficiencias físicas y psíquicas, nivel de ingresos, etc.

Resulta claro que esta fuente de error no puede ser neutralizada por medio de una investigación censal, ya que este tipo de error nada tiene que ver con la extensión de la muestra o población en estudio, sino que está directamente relacionada con la perturbación que la característica en estudio puede suscitar en el entrevistado ante una interrogación directa.

Cuando se investiga sobre temas tan delicados (sensitivos) que inducen a una fuerte resistencia por parte de los individuos en cuanto a respuestas francas, la Teoría de las Muestras aporta mejoras por vía de cuestionarios elaborados con mucha precaución, mayor entrenamiento de los entrevistadores tendiente a ganar la confianza de los entrevistados, reiteraciones y supervisión de las entrevistas, etc. Pero no siempre se logra eliminar el falseamiento o evasiva en la respuesta mediante estas medidas precautorias.

Desde hace varias décadas, los investigadores en el campo de las encuestas sociales se han esforzado en estudiar y crear nuevos procedimientos que eliminaran la inhibición del entrevistado.

En este sentido fue Stanley Warner el primero que elaboró un ingenioso método armado sobre la premisa de que los individuos estarían naturalmente mejor predispuestos a contestar con veracidad, si la técnica de formular las preguntas condujera a contestaciones no reveladoras de la pregunta contestada.

5.2 Modelos aleatorizados

En este punto se hará una reseña de algunos de los modelos aleatorizados, uno de los cuales será tratado en detalle.

5.2.1 Modelos Aleatorizados para poblaciones dicotómicas

Warner (1965) ideó el "Modelo Aleatorizado para la obtención de datos cualitativos en poblaciones dicotómicas con una característica íntima, sensitiva o estigmatizante". El mecanismo de aleatorización implica la formulación de dos proposiciones:

1) Pertenezco al grupo que posee la característica íntima.
2) No pertenezco al grupo que posee la característica íntima.

Una de estas dos proposiciones será contestada aleatoriamente por el entrevistado sin indicar a cuál de ellas se refiere.

En el caso de características inofensivas, inocuas, el estimador aleatorizado resulta ser menos eficiente que el estimador convencional, pero inmediatamente se torna más eficiente cuando la característica en estudio es de una intimidad tal, que induce a contestaciones falsas ante una interrogación directa (casos de intento de suicidio, drogadicción, alcoholismo, culpabilidad en accidentes automovilísticos, nivel de ingreso, deficiencias físicas o psíquicas, consumo de anfetaminas entre los estudiante universitarios, etc.).

5.2.2 Modelo Aleatorizado para poblaciones tricotómicas

Una ampliación del modelo de Warner fue el realizado por Abul Ela, Greemberg y Horvitz (1967) quienes lo trasladaron al caso de poblaciones multicotómicas y particularmente tricotómicas. No existe ningún basamento teórico para creer que este modelo sea más eficiente que el de Warner, sólo la creencia de suponer que los individuos se sentirían naturalmente más confiados si en vez de contestar aleatoriamente una de dos preguntas, lo hicieran con una de tres.

Para su aplicación se hace necesaria la extracción de dos muestras simples al azar independientes y no superpuestas. El principal mérito de este modelo reside en que sirvió de trampolín metodológico para pasar al modelo aleatorizado con una pregunta inocua.

5.2.3 Modelo Aleatorizado con una pregunta inocua

Simmons (1967) sostenía que la confianza de los entrevistados se vería incrementada si en vez de contestar aleatoriamente a una de dos preguntas, las dos referidas a la característica estigmatizante (modelo de Warner), lo hicieran con una de dos preguntas una de las cuales fuese inocua y totalmente desvinculada de la característica íntima. Esto cumple fundamentalmente el objetivo de eliminar la sospecha en los individuos de que el método pudiese encerrar una trampa del tipo "si sale cruz, gano y si sale cara pierdes".

Para su realización se hace necesaria la extracción de dos muestras simples al azar independientes y no superpuestas y la aplicación de la aleatorización en ambas muestras. El basamento teórico fue elaborado por Greemberg, Abul Ela, Simmons y Horvitz (1969).

5.2.4 Modelo Aleatorizado con una pregunta inocua, optimizado

En relación a una consideración sobre el mecanismo de aleatorización, que los autores hacen en el Modelo Aleatorizado con una pregunta inocua, Moors (1971) retoma el modelo y lo optimiza, aplicando la técnica de aleatorización a la primera muestra y un procedimiento de interrogación directa en la segunda muestra. Mediante este ingenioso artificio se logra que el modelo sea mucho más factible en cuanto a su aplicabilidad, más eficiente y a la vez menos costoso que el de Warner y que el modelo aleatorizado con una pregunta inocua.

5.2.5 Modelo Aleatorizado para la obtención de datos cuantitativos

Greemberg, Kuebler, Abernathy y Horvitz (1971) desarrollaron el basamento teórico del Modelo Aeatorizado para la obtención de datos cuantitativos. Este modelo conserva las mismas características que el Modelo Aleatorizado con una pregunta inocua en cuanto al mecanismo de aleatorización, pero en este caso la característica en estudio es susceptible de tomar valores cuantitativos.

5.2.6 Modelo lineal de respuestas aleatorizadas. Warner (1971) desarrolló un modelo lineal de respuestas aleatorizadas y demostró que todos los modelos aleatorizados existentes hasta entonces eran casos particulares del “Modelo lineal general”.

5.2.7 Modelo Aleatorizado con dos preguntas inocuas alternadas. Otro modelo para tratar de mejorar la eficiencia del estimador es el desarrollado por Folsom, Greemberg, Horvitz y Abernathy (1973). Este modelo se utiliza en el caso en que la proporción poblacional con la característica inocua es desconocida y por lo tanto se hace necesaria la extracción de dos muestras simples al azar independientes y no superpuestas. En este modelo se utilizan dos preguntas inocuas en forma alternada, además de la pregunta referida a la característica íntima en estudio.

5.2.8 Adaptación de un Modelo Cualitativo al Problema De La Estimación de Características Numéricas Continuas. Los modelos de respuestas aleatorizadas que se han expuesto cubren las situaciones de mayores posibilidades de aplicación. Estos se refieren a problemas que implican características cualitativas, siendo externos completamente a estos casos todas aquellas características de tipo métrico como podría ser por ejemplo la “distribución de la renta de las personas”. Gallese (1976) desarrolla una adaptación del modelo cualitativo con el objeto de obtener respuestas fidedignas sobre la distribución de los ingresos.

6 MODELO ALEATORIZADO PARA ESTIMAR PROPORCIONES EN POBLACIONES TRICOTOMICAS

El método aleatorizado para estimar proporciones en poblaciones humanas dicotómicas fue retomado y extendido al caso de poblaciones constituidas por tres o más grupos relacionados y mutuamente excluyentes por Abul Ela, Greemberg y Horvitz (1967) de la Universidad de North Caroline y el Research Triangle Institute.

6.1 Formulación del problema

Se tratará el caso en que los individuos pertenecen a uno sólo de los t ($t \geq 3$) grupos que la conforman, de los cuales al menos uno y a lo sumo t-1 poseen una característica estigmatizante o sensitiva. Para el caso en que t=3 se está en presencia de una población tricotómica: los individuos pertenecen al grupo A (que poseen la característica A) o al grupo B (que poseen la característica B) o al grupo C (que poseen la característica C), de los cuales al menos uno y a lo sumo dos son en diversos grados estigmatizantes. Se pretende por medio de una investigación muestral estimar las proporciones actuantes en la población en estudio.

6.2 Mecanismo de aleatorización

Para estimar la proporción de cada grupo en la población por medio de una muestra en el caso de utilizarse el procedimiento clásico, es decir requerimientos directos, se generaría un esquema trinomial. Para la aplicación de la técnica de respuestas aleatorizadas, en cambio, se requiere la extracción de dos muestras simples al azar con reposición, independientes y no superpuestas de tamaño n_1 y n_2 respectivamente. Se formulan tres proposiciones sobre una base probabilística. El entrevistado elige al azar una de las siguientes proposiciones:

1) pertenezco a A
2) pertenezco a B
3) pertenezco a C.

De estas tres proposiciones sólo una será afirmada o negada por el interrogado, según éste pertenezca o no al grupo que indique la pregunta elegida aleatoriamente, teniendo presente que la extracción de dos muestras implica la utilización de dos mecanismos diferentes entre sí, en cuanto a las probabilidades de elección de cada proposición. La composición de las urnas serán las mismas dentro de una misma muestra, pero serán distintas a las de la otra muestra.

Las probabilidades de elección generadas serán:

Pij: probabilidad de elegir la proposición j-ésima en la muestra n_i (i=1, 2) (j=1, 2, 3).

Evidentemente será:

$$\sum_{j=1}^{3} P_{ij} = 1 \qquad i = 1, 2.$$

6.3 Estimación aleatorizada de la proporción poblacional

Sea

π_1: proporción poblacional con la característica A

π_2: proporción poblacional con la característica B

π_3: proporción poblacional con la característica C

$$\sum_{j=1}^{3} \pi_j = 1$$

Se extraen dos muestras simples al azar independientes y no superpuestas de tamaño n_1 y n_2 a los efectos de hallar estimadores máximo verosímiles de las proporciones poblaciones. Se define

$$x_{ir} \begin{cases} =1 \text{ si el r-ésimo elemento de la } Mn_i \text{ es “si”} \\ =0 \text{ si el r-ésimo elemento de la } Mn_i \text{ es “no”} \end{cases}$$

i= 1, 2. r= 1, 2,..., n_i.

La probabilidad de que un individuo genérico de la primera muestra responda “si” es:

$$\lambda_1 = P(x_{1r} = 1) = \sum_{j=1}^{3} P_{1j} \; \pi_j$$

y como

$$\pi_3 = 1 - (\pi_1 + \pi_2)$$

resulta

$$\lambda_1 = (P_{11} - P_{13})\,\pi_1 + (P_{12} - P_{13})\,\pi_2 + P_{13} \qquad \text{(A)}$$

La probabilidad de que un individuo genérico de la primera muestra responda “no” es:

$$P(x_{1r} = 0) = 1 - \lambda_1$$

Simétricamente, para la segunda muestra se tiene:

$$\lambda_2 = P\,(x_{2r} = 1) = (P_{21} - P_{23})\,.\,\pi_1 + (P_{22} - P_{23})\,.\,\pi_2 + P_{23} \qquad \text{(B)}$$

y $P\,(x_{2r} = 0) = 1 - \lambda_2$

siendo condición que

K = (P11 – P13) (P22 – P23) – (P12 – P13) (P21 – P23) $\neq$ 0

Sea:

n_{i1}; cantidad de “si” en la i-ésima muestra.
$(n_i - n_{i1})$; cantidad de “no” en la i-ésima muestra
i = 1, 2.

la función de verosimilitud de la muestra será proporcional a

$$L = \lambda^{n_{11}}\,(1-\lambda_1)^{n_1 - n_{11}}\,\lambda_2^{\,n_{21}}\,(1-\lambda_2)^{n_2 - n_{21}}$$

logaritmando resulta

$$\log.\ L = n_{11}\log\lambda_1 + (n_1 - n_{11})\log(1-\lambda_1) + n_{21}\log\lambda_2 + (n_2 - n_{21})\log(1-\lambda_2)$$

Los estimadores máximos verosímiles de λ_1 y λ_2 serán los resultados del siguiente sistema de ecuaciones de verosimilitud.

$$\frac{\partial \log L}{\partial \pi_j} = 0\,, \qquad j = 1, 2.$$

$$\begin{cases} \dfrac{\partial \log L}{\partial \pi_1} = \dfrac{n_{11}}{\lambda_1}\dfrac{\partial \lambda_1}{\partial \pi_1} - \dfrac{n_1 - n_{11}}{1-\lambda_1}\dfrac{\partial \lambda_1}{\partial \pi_1} + \dfrac{n_{21}}{\lambda_2}\dfrac{\partial \lambda_2}{\partial \pi_1} - \dfrac{n_2 - n_{21}}{1-\lambda_2}\dfrac{\partial \lambda_2}{\partial \pi_1} = 0 \\[2ex] \dfrac{\partial \log L}{\partial \pi_2} = \dfrac{n_{11}}{\lambda_1}\dfrac{\partial \lambda_1}{\partial \pi_2} - \dfrac{n_1 - n_{11}}{1-\lambda_1}\dfrac{\partial \lambda_1}{\partial \pi_2} + \dfrac{n_{21}}{\lambda_2}\dfrac{\partial \lambda_2}{\partial \pi_2} - \dfrac{n_2 - n_{21}}{1-\lambda_2}\dfrac{\partial \lambda_2}{\partial \pi_2} = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial \log L}{\partial \pi_1} = \dfrac{n_{11}}{\lambda_1}(P_{11} - P_{13}) - \dfrac{n_1 - n_{11}}{1-\lambda_1}(P_{11} - P_{13}) + \dfrac{n_{21}}{\lambda_2}(P_{21} - P_{23}) - \dfrac{n_2 - n_{21}}{1-\lambda_2}(P_{21} - P_{23}) = 0 \\[2ex] \dfrac{\partial \log L}{\partial \pi_2} = \dfrac{n_{11}}{\lambda_1}(P_{12} - P_{13}) - \dfrac{n_1 - n_{11}}{1-\lambda_1}(P_{12} - P_{13}) + \dfrac{n_{21}}{\lambda_2}(P_{12} - P_{23}) - \dfrac{n_2 - n_{21}}{1-\lambda_2}(P_{22} - P_{23}) = 0 \end{cases}$$

$$\begin{array}{l} (P_{11} - P_{13})\,\dfrac{n_{11} - \lambda_1 n_1}{\lambda_1 (1-\lambda_1)} + (P_{21} - P_{23})\,\dfrac{n_{21} - \lambda_2 n_2}{\lambda_2 (1-\lambda_2)} = 0 \\[2ex] (P_{12} - P_{13})\,\dfrac{n_{11} - \lambda_1 n_1}{\lambda_1 (1-\lambda_1)} + (P_{22} - P_{23})\,\dfrac{n_{21} - \lambda_2 n_2}{\lambda_2 (1-\lambda_2)} = 0 \end{array} \qquad \text{(C)}$$

(C) constituye el sistema de ecuaciones de verosimilitud y sólo tendrá solución para

$$0 = \frac{n_{11} - \lambda_1\, n_1}{\lambda_1\,(1-\lambda_1)} = \frac{n_{21} - \lambda_2\, n_2}{\lambda_2\,(1-\lambda_2)} \qquad \text{(D)}$$

desde que

(P11 – P13) (P22 – P23) - (P12 – P13) (P21 – P23) ≠ 0

la solución (D) implica que

$$\hat{\lambda}_1 = \frac{n_{11}}{n_1} \qquad \text{y} \qquad \hat{\lambda}_2 = \frac{n_{21}}{n_2}$$

Si $\lambda_1 \neq 0$, $\lambda_2 \neq 0$

y $\lambda_1 \neq 1$, $\lambda_2 \neq 1$

sustituyendo en (A) y (B) se obtienen los estimadores máximo verosímiles de Π_1 y Π_2 y por ende de Π_3.

$$\text{(E)} \quad \begin{cases} (P_{11} - P_{13})\,\pi_1 + (P_{12} - P_{13})\,\pi_2 = \dfrac{n_{11}}{n_1} - P_{13} \\[2ex] (P_{21} - P_{23})\,\pi_1 + (P_{22} - P_{13})\,\pi_2 = \dfrac{n_{12}}{n_2} - P_{23} \end{cases}$$

de (E) resulta entonces

$$\hat{\pi}_1 = \left[\left(\frac{n_{11}}{n_1} - P_{13}\right)(P_{22} - P_{23}) - \left(\frac{n_{11}}{n_2} - P_{23}\right)(P_{12} - P_{13})\right] / k$$

$$\hat{\pi}_2 = \left[\left(\frac{n_{21}}{n_2} - P_{23}\right)(P_{11} - P_{13}) - \left(\frac{n_{11}}{n_1} - P_{13}\right)(P_{21} - P_{11})\right] / k$$

$$\hat{\pi}_3 = 1 - \left[\left(\frac{n_{11}}{n_1} - P_{13}\right)(P_{22} - P_{21}) - \left(\frac{n_{21}}{n_2} - P_{23}\right)(P_{12} - P_{11})\right] / k$$

Para determinar las propiedades de los estimadores se calculará su esperanza y su variancia.

Como $\frac{n_{i1}}{n_i}$ se distribuyen binomialmente con parámetros (λ_i, n_i) i = 1, 2., donde λ_1 y λ_2 son funciones de π_1 y π_2 como se define en (A) y (B), las esperanzas matemáticas de $\hat{\pi}_1, \hat{\pi}_2$ y $\hat{\pi}_3$ serán:

$$E(\hat{\pi}_1) = \frac{1}{k}\left[(\lambda_1 - P_{13})(P_{22} - P_{23}) - (\lambda_2 - P_{23})(P_{12} - P_{13})\right]$$

$$E(\hat{\pi}_2) = \frac{1}{k}\left[(\lambda_2 - P_{23})(P_{11} - P_{13}) - (\lambda_1 - P_{13})(P_{21} - P_{23})\right]$$

$$E(\hat{\pi}_3) = 1 - \left[E(\hat{\pi}_1) + E(\hat{\pi}_2)\right]$$

Sustituyendo λ_1 y λ_2 resulta

$$E(\hat{\pi}_1) = \frac{1}{k}\left\{\left[\pi_1 (P_{11} - P_{13}) + \pi_2 (P_{12} - P_{13})\right](P_{22} - P_{23}) - \left[\pi_1 (P_{21} - P_{23}) + \pi_2 (P_{22} - P_{23})\right](P_{12} - P_{13})\right\}$$

$$= \frac{1}{k} k\, \pi_1 \therefore E(\hat{\pi}_1) = \pi_1$$

$$E(\hat{\pi}_2) = \frac{1}{k}\left\{\left[\pi_1 (P_{21} - P_{23}) + \pi_2 (P_{22} - P_{13})\right](P_{11} - P_{13}) - \left[\pi_1 (P_{11} - P_{13}) + \pi_2 (P_{12} - P_{13})\right](P_{21} - P_{23})\right\}$$

$$= \frac{1}{k} k\, \pi_2 \therefore E(\hat{\pi}_2) = \pi_2$$

$$E(\hat{\pi}_3) = 1 - (\pi_1 + \pi_2) = \pi_3 \therefore$$

$$E(\hat{\pi}_3) = \pi_3$$

Es decir que el método aleatorizado conduce a estimadores no viciados en el caso de poblaciones tricotómicas y suponiendo respuestas veraces.

Se abordará el cálculo de las variancias de $\hat{\pi}_1, \hat{\pi}_2$ y $\hat{\pi}_3$ sabiendo que

$$\mathrm{Var}\left(\frac{n_{i1}}{n_i}\right) = \frac{\lambda_i (1-\lambda_i)}{n_i}, \qquad i = 1, 2$$

$$\mathrm{Var}(\hat{\pi}_1) = \frac{1}{k^2}\left[\frac{\lambda_1 (1-\lambda_1)}{n_1}(P_{22} - P_{23})^2 + \frac{\lambda_2 (1-\lambda_2)}{n_2}(P_{12} - P_{13})^2\right]$$

$$\mathrm{Var}(\hat{\pi}_2) = \frac{1}{k^2}\left[\frac{\lambda_1 (1-\lambda_1)}{n_1}(P_{21} - P_{13})^2 + \frac{\lambda_2 (1-\lambda_2)}{n_2}(P_{11} - P_{13})^2\right]$$

$$\mathrm{Var}(\hat{\pi}_3) = \frac{1}{k^2}\left[\frac{\lambda_1 (1-\lambda_1)}{n_1}(P_{22} - P_{21})^2 + \frac{\lambda_2 (1-\lambda_2)}{n_2}(P_{12} - P_{11})^2\right]$$

6.4. Determinación de los parámetros intervinientes en el modelo matemático

Las variancias de los estimadores en el modelo tricotómico convencional serían:

$$\mathrm{Var}\,(\hat{\pi}_j) = \frac{\pi_j\,(1-\pi_j)}{n}\ , \qquad j = 1, 2, 3;\ \ n = n_1 + n_2$$

Tanto en la obtención de esta fórmula como en el de la variancia aleatorizada se hizo el supuesto de que las preguntas son contestadas con veracidad.

La variancia del estimador convencional es sólo función de $n=n_1+n_2$ mientras que la variancia del estimador aleatorizado es función de los n_i (i=1,2) y de los Pij (j=1, 2,3.). Con el adecuado manipuleo de los Pij se podrá influir en el resultado de la variancia haciendo que ésta disminuya cuando los Pij maximicen su desvío con respecto a 1/3. Pero en la fijación de los Pij se habrá de tener en cuenta, además del efecto que produce sobre la variancia, la repercusión que va a tener en la aleatorización del mecanismo. Es decir que quizá haya que sacrificar un poco la precisión (teórica) del estimador para asegurar que el mecanismo conserve su aleatorización equilibrada en cuanto a su propósito primordial: inducir a respuestas francas por parte del interrogado. Si esto último no se cumple se echará por tierra todas las consideraciones teóricas puesto que éstas se hicieron suponiendo respuestas veraces.

En los puntos siguientes se podrá ver el efecto de las respuestas falseadas tanto en el modelo trinomial convencional como en el tricotómico aleatorizado.

6.5. Efecto de las respuestas falseadas

Dada la naturaleza de la característica en estudio, es dable pensar que tanto en la aplicación del procedimiento directo como en el del aleatorizado se filtrarán respuestas falseadas. Desde luego que estas filtraciones serán mucho menos significativas en el método aleatorizado, si logra su objetivo de incentivar la cooperación de los interrogados.

El siguiente es un ejemplo donde la característica estigmatizante o reservada es la C, la inocua la A, siendo la B una intermedia.

Si, por ejemplo, se quiere estimar la proporción de estudiantes, en una determinada Universidad o Facultad, que ingiere estimulantes químicos para aumentar su rendimiento; las proposiciones actuantes tanto en el modelo tricotómico aleatorizado como en el trinomial convencional son:

1) Pertenezco al grupo que nunca ingiere estimulantes químicos para estudiar. (característica A).
2) Pertenezco al grupo que a veces ingiere estimulantes químicos para estudiar. (característica B).
3) Pertenezco al grupo que habitualmente ingiere estimulantes químicos para estudiar. (característica C).

Las proporciones actuantes en la población serían:

π_1 : proporción poblacional con la característica A.

π_2 : proporción poblacional con la característica B.

π_3 : proporción poblacional con la característica C.

Introduciendo probabilidad de respuestas falseada en el modelo se demuestra que el método aleatorizado es más eficiente que el directo. Para profundizar sobre el tema ver Gallese (1976)

7 ADAPTACIÓN DE UN MODELO ALEATORIZADO CUALITATIVO AL PROBLEMA DE LA ESTIMACIÓN DE CARACTERÍSTICAS NUMÉRICAS CONTINUAS.

Los modelos de respuestas aleatorizadas que se han expuesto sucintamente, cubren las situaciones de mayores posibilidades de aplicación. Estos se refieren a problemas que implican características cualitativas, siendo externos completamente a estos casos todas aquellas características de tipo métrico como podría ser por ejemplo la “distribución de la renta de las personas”.

Si bien existe un modelo aleatorizado para la captación de datos cuantitativos, éste conduce a un estimador del promedio poblacional, pero de ningún modo describe la distribución de la variable numérica. Por ejemplo en el problema del Ingreso, el procedimiento de interrogación aleatorizada tratado a través de este modelo, proporciona un estimador del “promedio de ingreso por persona” pero no describe la “distribución de la renta de las personas”.

7.1 Formulación del problema. En lo que a Ingresos se refiere el problema fundamental es el de encontrar la “distribución de los ingresos de las personas”. Este a pesar de ser un problema de variable netamente cuantitativa, no puede atraparse con el “Modelo aleatorizado para la obtención de datos cuantitativos”.

Nos enfrentamos así con un caso de suma importancia práctica, en donde la característica básica es numérica y continua, vale decir que sale abiertamente del campo de todos los modelos que hemos presentado.

Sin embargo, este problema se lo puede transferir a un tratamiento de características cualitativas haciendo un parcelamiento en k categorías por medio de (k-1) divisiones en la escala de los ingresos.

La metodología para resolver este tipo de problema cuantitativo sometiéndolo a un tratamiento cualitativo se vislumbra en la misma Estadística Elemental. Por ejemplo, cuando se estudian distribuciones de variables continuas, los resultados obtenidos se clasifican según intervalos o clases dando lugar a las consabidas representaciones por histograma. Las cosas son así de tal naturaleza que lo que importa para los cálculos es el número de elementos que caen dentro de cada intervalo y no el valor particular que presentan dentro de cada intervalo. Las cosas son entonces de tal forma que, dos distribuciones diferentes según los valores individuales de sus elementos, pero tales que los mismos se distribuyen en cantidades iguales por cada uno de los intervalos, dan origen a histogramas similares.

El paso de la continuidad a un número de clases o intervalos contiguos es lo que siempre se presenta en toda consideración de matemática aplicada. Esto se debe a que no nos es posible seguir en los cálculos o análisis prácticos aplicados, las situaciones del continuo que se presentan en el caso de variables numéricas continuas.

7.2 Estimación de las proporciones poblacionales

En el problema de la renta, haciendo un fraccionamiento en la escala de los ingresos por medio de (k-1) divisiones, si π_j representa la proporción de personas pertenecientes a la categoría j-ésima (j= 1, 2,..., k), entonces será:

$$\sum_{j=1}^{k} \pi_j = 1$$

por lo tanto se habrán creado las condiciones necesarias que hacen factible la aplicación del “Modelo multicotómico aleatorizado”, para el caso en que k fuese igual a 3.
Se está pues en condiciones de formular el modelo que correspondería a una estimación de la “distribución de la renta” mediante (k-1) muestras independientes y no superpuestas.

Manteniendo para las π_j la significación de proporciones poblacionales (j= 1, 2,..., k), para las pij las probabilidades de que en la muestra i-ésima se enfrente la pregunta referida al intervalo j-ésimo y siendo λ_i, la probabilidad de una respuesta afirmativa en la muestra i-ésima (i= 1, 2,..., k-1), bajo el supuesto de respuestas veraces, se tiene:

$$\lambda_i = [P_{i,1} - P_{i,k}]\pi_1 + [P_{i,2} - P_{i,k}]\pi_2 + ... + [P_{i,k-1} - P_{i,k}] = \pi_{k-1} + P_{i,k}$$

Haciendo

$$[P_{i,1} - P_{i,k}] = \alpha_{i,1}, [P_{i,2} - P_{i,k}] = \alpha_{i,2}, ..., [P_{i,k-1} - P_{i,k}] = \alpha_{i,k-1}$$

y

$$[\lambda_i - P_{i,k}] = \alpha_{i,k}$$

se tiene

$$\alpha_{i,1}\,\pi_1 + \alpha_{i,2}\,\pi_2 + ... + \alpha_{i,k-1}\,\pi_{k-1} = \alpha_{i,k} \qquad \text{(H)}$$

i= 1, 2,..., k-1

Para k=4 el sistema de ecuaciones lineales (H), con 3 incógnitas en este caso, tiene por determinante de los coeficientes del sistema a

$$\Delta = \begin{vmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{vmatrix}$$

siendo (en virtud de la regla de Cramer) los numeradores, que divididos por Δ habrán de darnos los valores de π_1, π_2, π_3.

$$N_1 = \begin{vmatrix} \alpha_{1,4} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,4} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,4} & \alpha_{3,2} & \alpha_{3,3} \end{vmatrix}$$

$$N_2 = \begin{vmatrix} \alpha_{1,1} & \alpha_{1,4} & \alpha_{1,3} \\ \alpha_{2,1} & \alpha_{2,4} & \alpha_{2,3} \\ \alpha_{3,1} & \alpha_{3,4} & \alpha_{3,3} \end{vmatrix}$$

$$N_3 = \begin{vmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,4} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,4} \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,4} \end{vmatrix}$$

Desarrollando sucesivamente N1, N2 y N3 por los menores complementarios de los elementos $\alpha_{1,4}, \alpha_{2,4}, \alpha_{3,4}$ se llega a los siguientes resultados:

$$\pi_1 = \frac{\alpha_{1,4}\,\Delta_{1,1} - \alpha_{2,4}\,\Delta_{2,1} + \alpha_{3,4}\,\Delta_{3,1}}{\alpha_{1,1}\,\Delta_{1,1} - \alpha_{2,1}\,\Delta_{2,1} + \alpha_{3,1}\,\Delta_{3,1}}$$

$$\pi_2 = \frac{-\alpha_{1,4}\,\Delta_{1,2} + \alpha_{2,4}\,\Delta_{2,2} - \alpha_{3,4}\,\Delta_{3,2}}{\alpha_{1,2}\,\Delta_{1,2} - \alpha_{2,2}\,\Delta_{2,2} + \alpha_{3,2}\,\Delta_{3,2}} \qquad \text{(I)}$$

$$\pi_3 = \frac{\alpha_{1,4}\,\Delta_{1,3} - \alpha_{2,4}\,\Delta_{2,3} + \alpha_{3,4}\,\Delta_{3,3}}{\alpha_{1,3}\,\Delta_{1,3} - \alpha_{2,3}\,\Delta_{2,3} + \alpha_{3,3}\,\Delta_{3,3}}$$

Estos resultados corresponden al nivel idealizado del modelo. Recurriendo al método de máxima verosimilitud, los parámetros λ_i tendrán sus estimadores máximo verosímiles en sus correspondientes frecuencias relativas observadas, de suerte que las fórmulas (I) de los π_j deberá entenderse que se aplican a las estimaciones de las proporciones poblacionales.

El hecho de que se haya estimado la distribución (en este ejemplo de los Ingresos) mediante un número limitado de clases, en modo alguno puede extrañar a todo aquél que haya tratado, aún someramente, el problema del ingreso de las personas.

8 LA DISTRIBUCIÓN DE LOS INGRESOS. MODELOS.

La teoría de la distribución de los ingresos comprende dos aspectos diferentes pero interdependientes: uno considera el mecanismo remunerativo de los factores de la producción, el otro, los efectos de esa remuneración a través de la distribución de los ingresos entre los distintos grupos sociales.

8.1 Primera ley de Pareto

Wilfredo Pareto formuló la primera representación analítica de la curva de ingreso en cuanto a su distribución entre las personas. En uno de los capítulos de su célebre libro “Cours d’Economique Politique”, analiza la distribución de la renta en varios países y para diferentes épocas, concluyendo que la curva permanece invariada a través del tiempo y del espacio.

Si n(x) es la función de densidad de la renta x, y suponiendo que x pueda considerarse como variable continua; entonces n(x) dx representa el número de individuos con una renta entre x y x + dx. Si N(x) es el número de personas con un ingreso mayor o igual que x entonces:

$$N(x) = \int_x^{\infty} n(x)\,dx$$

La información estadística fiscal proporciona datos en cuanto a N(x).

Pareto concluyó que a partir de un nivel de renta adecuado, N(x) se expresa a través de la siguiente relación:

$$N(x) = \frac{A}{x^{\alpha}}$$

Siendo A y α constantes positivas. α representa la desigualdad de la distribución.

En todos los cálculos realizados por Pareto en distintos países y en diferentes años, los valores de α oscilan entre 1,89 y 1,24.

Pareto concluye: "Estos resultados son muy notables. Es absolutamente imposible admitir que éstos se deban solamente al azar. Hay ciertamente una causa que produce la tendencia de la renta conforme a una cierta curva. La forma de esta curva parece no depender más que levemente de las diferentes condiciones económicas de los países considerados".

La estimación de los parámetros de la primera ley de Pareto no ofrece dificultad dentro de los modelos de regresión lineal doble logarítmica.

8.2 Segunda ley de Pareto

En caso de que la transformación logarítmica no muestre los puntos ajustables por una función lineal entonces se usa la segunda ley de Pareto.

$$N(x) = \frac{A}{(x+a)^{\alpha}}$$

donde a es una constante positiva, negativa o nula.

8.2.1 Estimación de los parámetros de la segunda ley de Pareto

Logaritmando N(x) resulta:

$$\log N(x) = \log A - \alpha \lg (x + a)$$

como a no figura linealmente, no puede estimarse directamente por mínimos cuadrados.

Procedimiento de aproximaciones sucesivas.

El procedimiento analítico para estimar A, α y a en la segunda ley de Pareto es el de "Aproximaciones sucesivas".

Sea Y = log N(x) = log A - α log (x+a)

Y (x, H, α, a) = H - α log (x+a)

Sea h_0, α_0, y a_0 la primera aproximación obtenida de alguna manera y sean Δh, $\Delta\alpha$ y Δa las correcciones que deben ser aplicadas, o sea

$H = H_0 + \Delta H$

$\alpha = \alpha_0 + \Delta \alpha$

$a = a_0 + \Delta a$

Si se hace

$K = Y (x, H_0, \alpha_0, a_0) = H_0 - \alpha_0 \log (x+a_0)$

$A = Y'_H (x, H_0, \alpha_0, a_0) = 1$

$$B = Y'_{\alpha}\,(x, H_0, \alpha_0, a_0) = -\log(x + a)$$

$$C = Y'_{a}\,(x, H_0, \alpha_0, a_0) = -\frac{\alpha_0}{x + a_0}$$

donde las Y' denotan las derivadas parciales con respecto a las variables sub-indicadas.

Desarrollando Y como una serie de Taylor se tiene:

$$Y\,(x, H, \alpha, a) = K + A\,\Delta H + B\,\Delta\alpha + C\,\Delta a + \ldots$$

desde que K, A, B y C son funciones de x solamente, esta expresión es lineal en ΔH, $\Delta\alpha$ y Δa si se toma la aproximación hasta la primera derivada.

Aplicando mínimos cuadrados se obtienen el sistema de ecuaciones normales de Gauss:

$$\Delta H \sum A_i^2 + \Delta\alpha \sum B_i\ A_i + \Delta a \sum C_i\ A_i = \sum A_i\,[Y_i - K_i]$$

$$\Delta H \sum A_i\ B_i + \Delta\alpha \sum B_i^2 + \Delta a \sum C_i\ B_i = \sum B_i\,[Y_i - K_i]$$

$$\Delta H \sum A_i\ B_i + \Delta\alpha \sum B_i\ C_i + \Delta a \sum C_i^2 = \sum C_i\,[Y_i - K_i]$$

resolviendo este sistema de ecuaciones en ΔH, $\Delta\alpha$ y Δa se reconstruye

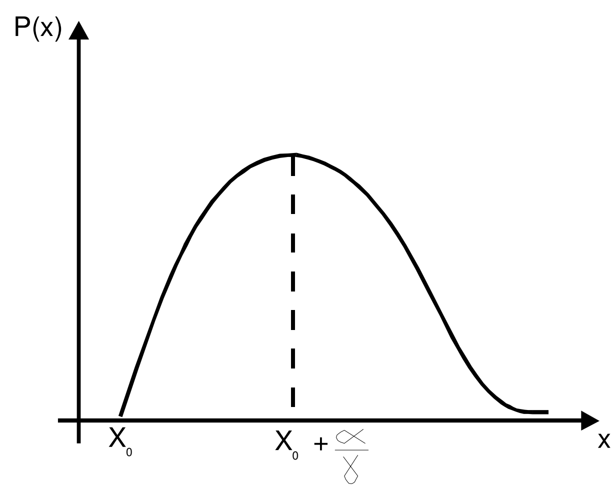
$$H = H_0 + \Delta H \qquad \alpha = \alpha_0 + \Delta\alpha \qquad a = a_0 + \Delta a$$

Tomando estos valores como segunda aproximación H_1, α_1, a_1, se reitera el procedimiento hasta que las correcciones sean despreciables.

8.3 La ecuación de Amoroso

El comportamiento de la distribución de la renta descripta por las leyes de Pareto dista mucho de ser el reflejo de la realidad económica ya que, las leyes de Pareto circunscriben el problema a los grupos superiores en cuanto a su ingreso. Uno de los tantos economistas que trata de dar una ecuación para la distribución de la renta superando los límites fiscales es Luigi Amoroso. La ecuación de Amoroso para la función de densidad en la distribución de los ingresos es del siguiente tipo:

$$P(x) = C(x - x_0)^{\alpha}\, e^{-\gamma\,(x - xo)} \qquad \alpha > 0,\ \gamma > 0$$



P(x) cae dentro del repertorio de frecuencias de Karl Pearson, siendo x_0 el mínimo necesario para la subsistencia, se determina a priori de realizar la investigación. El grave inconveniente para alimentar este modelo es la carencia de información estadística en cuanto a niveles de ingresos de las personas.

Si se desea conocer la distribución de los ingresos en una determinada región geográfica, el primer paso a dar en la investigación será en cuanto a la captación de información lo más fidedigna posible, ya sea a través del procedimiento de interrogaciones directas o a través de la "técnica de respuestas aleatorizadas". Una vez recabada la información, se la dispondrá en histogramas y se adoptará aquél que presente una forma más simple. Si responde a una forma campanular, unimodal, asimétrica a la derecha, podrá tenerse una primera impresión de que estos datos provienen de una población descripta por la ecuación de Amoroso.

8.3.1 Estimación de los parámetros por el método de los momentos

Haciendo $(x-x_0) = y$

$$p(y) = C\, y^{\alpha}\, e^{-\gamma y}$$

$$p'(y) = C\left[\alpha\, y^{\alpha-1}\, e^{-\gamma y} + y^{\alpha}\, (-\gamma)\, e^{-\gamma y}\right]$$

$$p'(y) = p(y)\left[\frac{\alpha}{y} - \gamma\right]$$

$$\frac{p'(y)}{p(y)} = \frac{\alpha - \gamma y}{y}$$

$$dx = dy$$

Los momentos teóricos son

$$\tilde{m}_s = \int p(x)\, x^s\, dx$$

$$\text{para } s = 0$$
$$\tilde{m}_0 = \int p(x)\, dx = 1$$

$$\text{para } s = 1$$
$$\tilde{m}_1 = \int p(x)\, x\, dx = E(x)$$

$$\text{para } s = 2$$
$$\tilde{m}_2 = \int p(x)\, x^2\, dx$$

$$\text{ahora bien}$$
$$\begin{aligned}\sigma^2 &= \int p(x)\,[x - E(x)]^2\, dx = \\ &= \int p(x)\, x^2\, dx + [E(x)]^2 \int p(x)\, dx - 2\,E(x) \int p(x)\, x\, dx = \\ &= \tilde{m}_2 + \tilde{m}_1^2 - 2\tilde{m}_1^2 \\ \sigma^2 &= \tilde{m}_2 - \tilde{m}_1^2\end{aligned}$$

Volviendo a la derivada logarítmica aplicando el procedimiento del método de los momentos resulta: Siendo m_s los momento de orden s, para s=0,1,2.

$$\frac{-[m_2 - m_1^2] + m_1^2}{m_2 - m_1^2} = \frac{m_1^2}{\sigma^2} - 1$$

$$\hat{\alpha} = \frac{m_1^2}{\sigma^2} - 1 \qquad \text{y como } \hat{\alpha} \text{ debe ser} \rangle\, 0$$

$$\text{deberá ser } m_1^2 \rangle\, \sigma^2$$
$$\text{de igual modo resultará}$$
$$\hat{\gamma} = \frac{m_1}{\sigma^2}$$

Entendiendo que tanto los momentos como la variancia serán los estimadores de los momentos y la variancia poblacional.

La condición necesaria para que la distribución se pueda ajustar con una curva del tipo Amoroso es que el momento de primer orden al cuadrado sea mayor que la variancia.

La constante C se determina por la condición de cierre

$$\int p(y)\,dy = 1$$

$$C \int_0^\infty y^\alpha e^{-\gamma y}\,dy = 1$$

resulta:

$$C = \gamma^{(\alpha+1)} / \Gamma(\alpha+1)$$

8.4 La función de Dagum

Otro ejemplo de función no lineal es la función propuesta por Dagum (1980) para modelar la distribución de los ingresos.

$$F(x) = \alpha + (1-\alpha)\left(1+\lambda x^{-\delta}\right)^{-\beta} \qquad x \geq 0,\quad 0 \leq \alpha \leq 1,\quad \beta > 0,\quad \lambda > 0,\quad \delta > 0$$

Donde *x* es el ingreso y $F(x) = P(X \leq x)$ es su función de distribución acumulada.

La correspondiente función de densidad viene dada por:

$$f(x,\beta,\lambda,\delta) = \begin{cases} 0 & x < 0 \\ \alpha & x = 0 \\ (1-\alpha)\beta\lambda\delta x^{-\delta-1}\left(1+\lambda x^{-\delta}\right)^{-\beta-1} & x > 0 \end{cases}$$

En la mayoría de los casos se supone $\alpha = 0$, por lo tanto resulta:

$$f(x,\beta,\lambda,\delta) = \beta\lambda\delta x^{-\delta-1}\left(1+\lambda x^{-\delta}\right)^{-\beta-1} \qquad x > 0$$

A diferencia de la distribución de Pareto, no existe ninguna transformación que lleve esta expresión a una función lineal en los parámetros. Por lo tanto, no es factible ajustar un modelo de regresión lineal y para la estimación de los parámetros se debe recurrir a métodos de ajuste de funciones no lineales.

8.4.1 Estimación de parámetros de la función de Dagum

Uno de los procedimientos analíticos para estimar los parámetros α, β, λ y δ es el de "Aproximaciones sucesivas", también llamado "Método iterativo de linearización".

En primera instancia, se expresa la función no lineal como una expansión de series de Taylor, alrededor de algún conjunto inicial de valores de los coeficientes: α_0, β_0, λ_0 y δ_0. La siguiente ecuación proporciona una aproximación lineal a la función cuyos parámetros se desean estimar.

$$F(x,\alpha,\beta,\lambda,\delta) = F(x,\beta_0,\lambda_0,\delta_0) + (\alpha-\alpha_0)\left.\frac{\partial F}{\partial \alpha}\right|_{\alpha=\alpha_0} + (\beta-\beta_0)\left.\frac{\partial F}{\partial \beta}\right|_{\beta=\beta_0} + (\lambda-\lambda_0)\left.\frac{\partial F}{\partial \lambda}\right|_{\lambda=\lambda_0} + (\delta-\delta_0)\left.\frac{\partial F}{\partial \delta}\right|_{\delta=\delta_0} + \varepsilon \quad (*)$$

donde

$$\frac{\partial F}{\partial \alpha}=1-\left(1+\lambda x^{-\delta}\right)^{-\beta}$$

$$\frac{\partial F}{\partial \beta}=-(1-\alpha)\left(1+\lambda x^{-\delta}\right)^{-\beta}\ln\left(1+\lambda x^{-\delta}\right)$$

$$\frac{\partial F}{\partial \lambda}=-(1-\alpha)\,\beta\left(1+\lambda x^{-\delta}\right)^{-\beta-1}x^{-\delta}$$

$$\frac{\partial F}{\partial \delta}=(1-\alpha)\,\beta\lambda x^{-\delta}\left(1+\lambda x^{-\delta}\right)^{-\beta-1}\ln x$$

Llamando $V_0(x)=\left.\frac{\partial F}{\partial \alpha}\right|_{\alpha=\alpha_0}$ $W_0(x)=\left.\frac{\partial F}{\partial \beta}\right|_{\beta=\beta_0}$ $Y_0(x)=\left.\frac{\partial F}{\partial \lambda}\right|_{\lambda=\lambda_0}$ $Z_0(x)=\left.\frac{\partial F}{\partial \delta}\right|_{\delta=\delta_0}$

La expresión (*) resulta:

$$F(x,\alpha,\beta,\lambda,\delta)=F(x,\alpha_0,\beta_0,\lambda_0,\delta_0)+(\alpha-\alpha_0)V_0(x)+(\beta-\beta_0)W_0(x)+(\lambda-\lambda_0)Y_0(x)+(\delta-\delta_0)Z_0(x)+\varepsilon$$

$$F(x,\alpha,\beta,\lambda,\delta)-F(x,\alpha_0,\beta_0,\lambda_0,\delta_0)+\alpha_0V_0(x)+\beta_0W_0(x)+\lambda_0Y_0(x)+\delta_0Z_0(x)=\alpha V_0(x)+\beta W_0(x)+\lambda Y_0(x)+\delta Z_0(x)+\varepsilon$$

Siendo $u(x)=F(x,\alpha,\beta,\lambda,\delta)-F(x,\alpha_0,\beta_0,\lambda_0,\delta_0)+\alpha_0V_0(x)+\beta_0W_0(x)+\lambda_0Y_0(x)+\delta_0Z_0(x)$

la ecuación anterior resulta:

$$u(x)=\alpha V_0(x)+\beta W_0(x)+\lambda Y_0(x)+\delta Z_0(x)+\varepsilon$$

Esta ecuación es lineal en los parámetros α, β, λ y δ y puede expresarse matricialmente de la forma: $\mathbf{U}=\mathbf{X\theta}+\boldsymbol{\varepsilon}$

donde:

$$\mathbf{U}=\begin{bmatrix}u_1\\ \vdots\\ u_n\end{bmatrix}\quad \mathbf{X_0}=\begin{bmatrix}V_0(1) & W_0(1) & Y_0(1) & Z_0(1)\\ \vdots & \vdots & \vdots & \vdots\\ V_0(n) & W_0(n) & Y_0(n) & Z_0(n)\end{bmatrix}\quad \boldsymbol{\theta}=\begin{bmatrix}\alpha\\ \beta\\ \lambda\\ \delta\end{bmatrix}\quad \boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{bmatrix}$$

De este modo, los coeficientes pueden estimarse a través del método de mínimos cuadrados ordinarios, como se observa a continuación:

$$\hat{\boldsymbol{\theta}}_1=\begin{bmatrix}\alpha_1\\ \beta_1\\ \lambda_1\\ \delta_1\end{bmatrix}=\left(\mathbf{X}_0'\mathbf{X}_0\right)^{-1}\mathbf{X}_0'\mathbf{U}=\boldsymbol{\Phi}\left(\boldsymbol{\theta}_0\right)$$

Tomando los valores $\alpha_1,\beta_1,\lambda_1$ y δ_1 como un conjunto nuevo de estimaciones iniciales, se reitera el procedimiento, desarrollando la serie de Taylor en estos nuevos valores:

$$F(x,\alpha,\beta,\lambda,\delta)-F(x,\alpha_1,\beta_1,\lambda_1,\delta_1)+\alpha_1V_1(x)+\beta_1W_1(x)+\lambda_1Y_1(x)+\delta_1Z_1(x)=\alpha V_1(x)+\beta W_1(x)+\lambda Y_1(x)+\delta Z_1(x)+\varepsilon$$

Aplicando nuevamente el método de mínimos cuadrados ordinarios, se obtiene un nuevo conjunto de estimaciones de los coeficientes: $\hat{\boldsymbol{\theta}}_2=\boldsymbol{\Phi}\left(\hat{\boldsymbol{\theta}}_1\right)$. Se repite el procedimiento reiteradas veces:

$$\hat{\boldsymbol{\theta}}_3=\boldsymbol{\Phi}\left(\hat{\boldsymbol{\theta}}_2\right)$$
$$\vdots$$
$$\hat{\boldsymbol{\theta}}_n=\boldsymbol{\Phi}\left(\hat{\boldsymbol{\theta}}_{n=1}\right)$$

hasta que se alcanza la convergencia, es decir, cuando:

$$\left|\frac{\alpha_{j+1}-\alpha_j}{\alpha_j}\right|<\xi \qquad \left|\frac{\beta_{j+1}-\beta_j}{\beta_j}\right|<\xi \qquad \left|\frac{\lambda_{j+1}-\lambda_j}{\lambda_j}\right|<\xi \qquad \left|\frac{\delta_{j+1}-\delta_j}{\delta_j}\right|<\xi$$

donde ξ es un número pequeño cuya elección depende en parte del costo de cálculo. No obstante, no se tiene garantía de que este proceso iterativo convergerá en la estimación máxima verosímil de los parámetros. Ya que, por ejemplo, puede converger en un mínimo local de la función de suma de cuadrados de los errores.

En la práctica se aconseja reiterar la estimación variando el conjunto de conjeturas iniciales.

Un método alternativo implica una variación en el método antes descrito. En lugar de usar las estimaciones sucesivas resultantes de cada linealización, las estimaciones son calculadas a partir de:

$$\boldsymbol{\theta}_{j+1}=\boldsymbol{\theta}_j+\alpha(\hat{\boldsymbol{\theta}}_{j+1}-\boldsymbol{\theta}_j) \qquad |\alpha|<1$$

donde $\hat{\boldsymbol{\theta}}_{j+1}$ es la estimación de mínimos cuadrados de la $(j+1)$ ésima iteración.

Según el método de Newton Raphson o método de las tangentes:

$$\alpha=\frac{1}{1-\boldsymbol{\Phi}\left(\hat{\boldsymbol{\theta}}_j\right)}$$

por lo tanto:

$$\boldsymbol{\theta}_{j+1}=\boldsymbol{\theta}_j+\frac{\hat{\boldsymbol{\theta}}_{j+1}-\boldsymbol{\theta}_j}{1-\boldsymbol{\Phi}'\left(\hat{\boldsymbol{\theta}}_j\right)}$$

Los estimadores obtenidos a través del procedimiento de mínimos cuadrados tienen una distribución asintótica normal[5], es decir:

$$\hat{\boldsymbol{\theta}} \underset{D}{\rightarrow} N\left[\boldsymbol{\theta}; \frac{\sigma^2}{n}.\mathbf{Q}^{-1}\right]$$

donde:

$$\mathbf{Q} = P\lim_{n\to\infty}\frac{1}{n}.\mathbf{X'X}$$

y la estimación muestral de la matriz de variancias y covariancias asintótica es:

$$\mathbf{V\hat{a}r}(\hat{\boldsymbol{\theta}}) = \hat{\sigma}^2\left(\mathbf{X'X}\right)^{-1}$$

donde **X** es la matriz valorizada en $\hat{\boldsymbol{\theta}}$.

9 Conclusiones y reflexiones

Siguiendo las recomendaciones internacionales sobre fomentar la interacción entre la producción de datos, la investigación científica y la enseñanza, en este trabajo los investigadores docentes, actuando como nexo, les dan participación a los alumnos y desarrollan un método para captar información garantizando la privacidad de la misma, destinado a los productores de datos para su aplicación.

Esta pretensión se concreta incentivando a los alumnos para que se involucren y participen, capacitándolos para ello, prestándole asistencia tutorial permanente, fomentando el trabajo autónomo y colaborativo y dándoles participación en los proyectos de investigación. De hecho, hay cuatro ex alumnos y tres alumnos participando en esta empresa.

Resultado de fomentar este trabajo colaborativo es que se enfatiza el contexto del problema planteado:

Dentro de los Objetivos de Desarrollo del Milenio. Para mejorar el monitoreo del cumplimiento de las metas del primer objetivo.

Dentro de un esquema simplificado de una economía nacional. Para ubicar el tema dentro de la teoría económica.

Dentro de la investigación aplicada. El tema desarrollado se enmarca dentro de la dimensión del ciclo investigativo y de los tipos de pensamiento estadístico.

El vínculo entre la producción de los datos de calidad y el modelado de excelencia se establece por medio del ajuste de funciones.

Se enfatiza sobre la necesidad de datos de calidad presentando un recorrido por los distintos modelos que refieren a la técnica de interrogaciones aleatorizadas.

Existe una razón poderosa para haber presentado este método de captación de la información en el área de los argumentos individuales íntimos y aún empresas con coordenadas informativas reservadas Ésta es no marginar la vida entera con sus contenidos de prejuicios

[5] Ver William H. Greene. Página 395

o temores de los individuos para planear subsiguientes reformas sociales, o bien superar las vallas restrictivas de diversos intereses de personas o empresas para arbitrar soluciones de justicia equitativa.

La conclusión fundamental es que, manteniendo el supremo postulado de conservar el anonimato de las declaraciones originales, se puede adquirir información en el mundo íntimo de las características individuales.

Se sabe bien que el Cálculo de las Probabilidades en punto a la moderna Teoría de las Muestras se ha ubicado en la confección del diseño de las mismas. Ahora bien, en la subsiguiente Teoría de las Interrogaciones Aleatorizadas, el Cálculo de las Probabilidades se viene a ubicar en el mismo proceso donde se realiza la captación original de información estadística, conservando el más absoluto anonimato de las declaraciones de cada informante.

Es el momento de recalcar que el genio creador de esta técnica fue el estadístico Stanley L. Warner y que sobre su descubrimiento se basaron los ulteriores modelos de otros autores. Actualmente métodos similares se utilizan para preservar la privacidad de la información. A las múltiples macro matrices de datos que circulan por el cyber espacio, se los “perturba” con bombardeos probabilísticos de suerte que no sea posible reconstruir la información brindada individualmente. Sin embargo, mantienen su potencial estadístico.

Se deja en manos de los jóvenes y entusiastas investigadores la tarea de ahondar en este terreno, la toma de información preservando la privacidad, dado que cubre un amplio espectro de aplicación. En la actualidad, la privacidad se puede lograr ya sea desde el mecanismo mismo de la toma de la información, asegurando al informante que su respuesta individual no es detectable ni siquiera por el propio entrevistador, ya sea “disfrazando aleatoriamente” los datos (ya recabados) de las múltiples macro matrices del cyber espacio.

REFERENCIAS BIBLIOGRÁFICAS

Abul Ela, Abdel Latif; Bernard G. Greemberg y Horvitz, Daniel G. (1967). A Multi-Proportions Randomized Response Model. *Journal of the American Statistical Association,* 62. September 1967. 990-1008.

Abul Ela, Abdel Latif; Simmons, Walt R. (1969). The Unrelated Question Randomized Response Model. Theorical Framework. *Journal of the American Statistical Association*. 64 .June 1969, 520, 39.

Dagum, C. (1980). Inecuality Measures Between Income Distributions with applications. *Econometrica.* 48. 7. Pág. 1791-1814

Dinur, I. y Nissim, K. (2003). Revealing information while Preserving Privacy. *<http://delivery.acm.org/10.1145/780000/773173/p202-dinur.pdf?key1=773173&key2=5192773611&coll=GUIDE&dl=portal,ACM&CFID=11111111&CFTOKEN=2222222>.* [15/10/2005)

Horvitz, Daniel G.; Sha, B. V. y Simmons, Walt R. (1967). The Unrelated Question Randomized Response Model. *Proceedings of Social Statistics Section*, Washington, D.C.: American Statistical Association.

Gallese, E. (1976) *Nuevos métodos probabilísticos para la captación de información estadística*. Tesis doctoral. Universidad Nacional de Rosario.

Greemberg, Bernard G.; Kuebler, Roy R. Jr.; Abernathy, James R. y Horvitz, Daniel G. (1971). Application of the Randomized Response Thecnique in Obtaining Quantitative Data. *Journal of the American Statistical Association*. 66 (June 1971) 243-50.

Moors, J.J.A. (1971). Optimization of the Unrelated Question Randomized Response Model. *Journal of the American Statistical Association*. 66 .September 1971. 627-29.

Warner, Stanley L. (1965). Randomized Response: A Survey Tecnique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*. 60. March 1965. 63-9.

Warner, Stanley L. (1971). The Linear Randomized Response Model. *Journal of the American Statistical Association*. 66 .December 1971. 884-88.

Wild, C. J. y Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry *International Statistical Review* 1999. 67, 3. Pág. 223-266.

Presidencia de la Nación (2003) *Objetivos de Desarrollo del Milenio. Argentina. La oportunidad para su reencuentro.*

BIBIOGRAFÍA COMPLEMENTARIA

Batholomew, D (1995). What is Statistics? *J. R. statistic Soc. A*, **158**, Part 1, 1-20.

Du, W. y Zhan, Z. (2003). Using Randomized Response Techniques for Privacy-Preserving Data Mining.
<http://sai.syr.edu/facultypapers/Randomized%20Response%20Techniques.pdf>. [15/10/2005].

Dwork, C. ,Mc Sehrry, F., Nissim, K. y Smith, a. (2004). Calibrating Noise to Sensitivity in Private Data Analysis.
< http://research.microsoft.com/research/sv/DatabasePrivacy/sensitivity.pdf> [12/09/2006]

Folsom. Ralph E.; Greemberg, Bernard G.; Horvitz, Daniel G. y Abernathy, James R. (1973). The Two Alternate Question Randomized Response Model for Human Surveys. *Journal of the American Statistical Association*. 68 .September 1973. 525-30.

Gallese, Elda. (1973) Técnica de respuestas aleatorizadas para la obtención de datos cuantitativos. *Revista Facultad de Ciencias Económicas de la U. N. de Cuyo*. 73/75 Enero-Diciembre 1973.

Greemberg, Bernard G.; Abernathy, James R. y Horvitz, Daniel G. (1970). A New Surveys Tecnique and Its Application in the Field of Public Health, Milbank *Memorial Fund Quarterly*. 48. Part 2 October 1970. 39-55.

Greene, W. H. (1996). *Análisis Econométrico* Tercera Edición. Prentice Hall.

Hoaglin, D. C. y Moore, D. S (1992). Editors: *Perspectives on Contemporary Statistics*. Mathematical Association of America. MAA Notes Number 22. 1992.

Kingstone, Jorge. (1953). *La Desigualdad en la Distribución de la Renta*. Biblioteca Interamericana de Estadística (B.I.E.T.A.). Traducción Dora Musciatti.

Kolmogorov, A. N. y Fomín, S. V. (1975). *Elementos de la teoría de funciones y del análisis funcional*. Editorial MIR. Moscú.

Luthe, R.; Olivera, A. y Schutz, F. (1986). *Métodos Numéricos*; Editorial Limusa; Primera edición; México; 443p.

Moore, D. S. (1997). New pedagogy and new contents: the case of statistics. *International Statistical Review*. 65, 123-166.

Pindyck, R. S. y Rubinfeld, D. L.(2000). *Econometría, modelos y pronósticos*. Mc Graw Hill. Cuarta edición. México.

Rietz, H. L. (1924). *Handbook of Mathematical Statistics*. Houghton Mifflin Company.

Walter, G. y Das Neves, F. (2006). Business Intelligence y Data Mining. *Snoop files*.