
Estimación por muestreo del índice de Gini para las localidades de Bogotá D.C. usando funciones en R¹

Gini index sampling estimation for localities in Bogotá D.C. using R functions

José Andrés Flórez^a
andres.florez@colfuturo.org

Hugo Andrés Gutiérrez^b
hugogutierrez@usantotomas.edu.co

José Fernando Zea^c
josezea@usantotomas.edu.co

Resumen

En este tiempo es de gran importancia los datos que provienen de las medidas de desigualdad, en economía y las ciencias sociales. En este artículo se utiliza esta información para calcular el índice de Gini, el cual se basa en los ingresos de una población, más específicamente en Bogotá D.C. para sus diecinueve localidades. Se utiliza la Encuesta Multipropósito para Bogotá Distrito Capital - EMB - 2011 junto con sus módulos específicos usando funciones en el software estadístico R para estimar (utilizando pesos de muestreo) el índice de Gini por cada localidad definiendo las variables que incluyen la información acerca de los ingresos por hogar.

Palabras clave: encuesta multipropósito de Bogotá D.C., factores de expansión, índice de Gini, medidas de desigualdad, R.

Abstract

In this time it is important the data with respect to inequality measures in economics and social science. In this paper we use this information to calculate the Gini index. This index uses the information about the income in a population, specifically in Bogotá D.C. for their 19 localities. We use the multipurpose survey

¹Flórez, J. A., Gutiérrez, H. A., Zea, J. F. (2015). Estimación por muestreo del índice de Gini para las localidades de Bogotá D.C. usando funciones en R. *Comunicaciones en Estadística*, **8**(1), 59-79.

^aAsistente Gestión Becas COLCIENCIAS y Estadísticas. Colfuturo. Colombia

^bDocente. Facultad de Estadística. Universidad Santo Tomás. Colombia

^cDocente. Facultad de Estadística. Universidad Santo Tomás. Colombia

for Bogotá D.C. in 2011 and use the modules for this analysis in each locality defining the variables that included information about incomes per household.

Keywords: Bogotá D.C. multipurpose survey, Gini index, inequality measures, R, sample weights.

1. Índice de Gini

Como se muestra en Chotikapanich (2010) la Curva de Lorenz cumple con un rol de descripción frente a las medidas de desigualdad. Da la proporción de la población con ingresos y compara las distribuciones del mismo. El índice de Gini se encuentra entre la línea hipotética de igualdad y la Curva de Lorenz.

El índice de Gini mide el grado de distribución de los ingresos o el consumo entre los individuos u hogares dentro de una economía perfectamente distribuida, en adición es posible analizar la desigualdad de la riqueza. El índice de Gini se encuentra en un intervalo entre 0 y 1, al momento de tener una distribución completamente equitativa de los ingresos tomaría un valor de cero, donde la población tiene una distribución igual del total de los ingresos, en caso contrario de una distribución desigual perfecta tomará un valor de 1.

Gini es igual a dos veces el área entre la curva de Lorenz y la diagonal.

$$G = 1 - \frac{2}{N-1} \left(N - \frac{\sum_{i=1}^N ix_i}{\sum_{i=1}^N x_i} \right) \quad (1)$$

En este trabajo generalizamos la forma expandida del índice de Gini presentado por Mark (1999).

Siendo \hat{N} el tamaño poblacional estimado, w_i el peso de muestreo (inverso de la probabilidad de inclusión o peso de calibración) y x_i el ingreso del i -ésimo individuo, se propone la siguiente expresión como estimador del Gini poblacional:

$$\hat{G} = \sum_{k=2}^n \left[\sum_{i=2}^k \frac{w_i}{\hat{N}} x_i \sum_{i=1}^{k-1} \frac{w_i}{\hat{N}} x_i \right] - \left[\sum_{i=1}^{k-1} \frac{w_i}{\hat{N}} x_i \sum_{i=2}^k \frac{w_i}{\hat{N}} x_i \right] \quad (2)$$

2. Encuesta Multipropósito para Bogotá Distrito Capital - EMB - 2011

Una base importante de nuestro estudio para el cual usamos el índice de Gini, es en el uso de la Encuesta Multipropósito para Bogotá D.C., 2011, esto con el propósito de identificar la relación entre las diecinueve localidades de Bogotá D.C.. Debemos tener en cuenta estudios anteriores del Departamento Nacional de Estadísticas

(DANE), la Secretaría Distrital de Planeación (SDP), Gobierno Nacional y las encuestas del Gobierno Distrital, ya que es una forma de conocer cómo es la calidad de vida en nuestra ciudad, en las diecinueve áreas urbanas.

A lo largo de varios años, 1991, 1993, 2003 y 2007, el DANE realizó una encuesta en la que se media la calidad de la vida en las personas, en 2004 el Centro de Investigaciones para el Desarrollo (CID) de la Universidad Nacional de Colombia llevó a cabo el estudio sobre la capacidad de pago, recibiendo por nombre: Encuesta sobre la equidad en las tarifas de servicios: impacto en la capacidad de pago de los hogares en Bogotá D.C. Luego, se realizó la misma encuesta representativa para el total nacional y el tema se ha mantenido vigente a lo largo de sus diferentes versiones.

Al conocer el comportamiento de las políticas distritales aplicadas, con la Encuesta Multipropósito para Bogotá D.C., 2011 mostraremos el impacto socioeconómico de la misma, junto con un enlace entre las encuestas de calidad de vida y la habilidad de pago manteniendo los niveles de desagregación por localidad y estrato. Los hogares serán la unidad de observación y los individuos quienes lo componen, también las casas que se encuentren dentro o cerca de la misma en la zona.

Queremos nombrar los principales propósitos de la Encuesta Multipropósito y generar la información para futuras interpretaciones, también evidenciar que (Handcock & Morris 1999) el índice de Gini es una medida estándar de desigualdad que satisface el principio de invariancia de escala; en algunos casos es importante preservar la diferencia absoluta del dólar para posibles diferencias en las comparaciones.

Dentro de los objetivos se pretende tener información acerca de la calidad de vida de las personas que se encuentran en la ciudad, esto hace referencia a la conducta por unas variables que se explican en la encuesta que fueron aplicadas a las localidades de Bogotá D.C..

- El conocimiento de los aspectos importantes que mide localidad del ambiente en el hogar o lugar de residencia, uno de estos aspectos es la contaminación visual conocida como acinamiento o sobrepoblación.
- Tener las estadísticas como base importante de los indicadores para reconocer los problemas de las personas que residen en la parte urbana de Bogotá D.C..
- Analizar los problemas urbanos para los habitantes de Bogotá D.C. en aspectos como la educación, salud, empleo (caracterización y justificación), conocimientos de las nuevas tecnologías de la información (TICS), si estas reciben el uso adecuado o si no tienen acceso, al igual que la percepción de seguridad, su participación en las redes sociales y finalmente su satisfacción con la vida urbana que llevan.
- Generar información para la construcción de indicadores sociales por sectores, localidades y diferentes estratos de la ciudad; hay un claro interés en conocer la cantidad de ingresos que las personas tienen y el déficit de vivienda para aquellos que no tienen acceso a coberturas sociales.

La encuesta ha tenido la influencia de conceptos básicos que hacen referencia a la población que vive en Bogotá D.C. con características muy detalladas, por lo cual es importante resaltar que la principal temática y sus componentes, consiste en obtener bastante información en lo que respecta a la calidad de vida urbana, capacidad de pago y los demás servicios utilizados en los hogares, por este motivo describiremos cada uno de los módulos que componen la encuesta.

- Características de los hogares y su ambiente, determinado por la construcción interna y externa que tienen, concreto, madera, tipos de pisos, esto nos brinda la capacidad de entender problemas con inestabilidad de la tierra por ubicarse en sectores donde no se debe construir.
- Si la casa es propia o es en arriendo, si es habitada por uno o más hogares, el pago de impuestos, si reciben subsidio o si han recibido alguna ayuda social. Frente a los servicios públicos como electricidad, agua, gas, recolección de basuras y la frecuencia con que acceden a este servicio.
- También encontramos variables demográficas como el sexo, la edad, la relación que se tiene en el hogar, el estado, si a sufrido migración o si es desplazado por algún tipo de violencia y aquellas personas que viven en el hogar.
- Educación, nivel de alfabetización, asistencia a los colegios o escuelas públicas destinadas por el distrito, también aquellas razones por las cuales no asisten a los colegios o escuelas por recursos, distancias, porque no tienen una persona que cuide en el camino al colegio, en los beneficios y defectos del sistema. Es importante ver qué actividades lúdicas realizan, por ejemplo practicar un deporte o si hacen trabajo infantil, el cual no les permite desarrollar sus habilidades de aprendizaje.
- El uso de las tecnologías de la información, las cuales tienen grandes ventajas para las personas con dificultad de interacción, la facilidad para acceder a ellas, si se tiene que dirigir a algún lugar o si puede acceder desde su hogar.
- La percepción que tienen las personas acerca de las condiciones de vida y describir qué está pasando en su ciudad, también lo que pasa en sus hogares, para la familia y como cabeza de hogar, la frecuencia en que ocurren problemas de crímenes, robos, problemas de tráfico y lo relacionado en años anteriores.
- El gasto familiar en alimentos, bienes y servicios, periodicidad de compra, dependiendo el tipo de bien que se está adquiriendo.

Para la encuesta es significativa la cobertura geográfica, que tendremos en cuenta, dado que tomaremos diecinueve de las veinte localidades. Sumapaz es la localidad número veinte, sin embargo no se incluye en el estudio debido a que es rural, y nuestra unidad de observación se encuentra en el área urbana. El universo de la Encuesta Multipropósito para Bogotá D.C., 2011 está compuesta por los hogares

y la población civil no institucional existente en 2011 en el distrito urbano de la capital.

Así el universo objetivo se encuentra compuesto por hogares y población civil no institucional para el año 2011 en el área urbana, con exclusión de las personas en las áreas rurales de la ciudad en Usaquén, Chicó, Santa Fe, San Cristobal, Usme, Suba, Sumapaz y Ciudad Bolívar. Cárceles o prisiones, centros de rehabilitación, orfanatos u hogares de paso, hogares geriátricos, seminaristas y monasterios, pensiones de estudiantes, cuarteles o guarniciones, estaciones de policía, campos de trabajo, centros de acogida para desplazados y rehabilitados, y centros de agricultura.

La información fue proporcionada por el Departamento Nacional de Estadística (DANE) y la Secretaría Distrital de Planeación de Bogotá, (SDP). Los investigadores primarios fueron Garavito y Villamizar.

La encuesta se compone de varios módulos, que se encuentran diseñados para conocer la forma de vida de cada persona, su condición y la diferencia frente a otras personas. Procedemos a dar una descripción de cada uno, para una mayor claridad de referencias.

La encuesta se realizó bajo un muestreo probabilístico, estratificado y de conglomerados urbanos. Por localidad y estrato el tamaño de muestra por segmentos es de 1.860 y el tamaño de muestra por predios es de 14.779. La EMB de 2011 es la primera encuesta que congrega información para cambios en niveles de pobreza y vida urbana, entre otros.

A. Identificación

Contiene la información acerca de la ubicación, la industria, la sección, el número del hogar, total de integrantes del hogar, estrato y el total de las personas que viven en el hogar. Es de gran importancia que las respuestas sean verdaderas, a fin de realizar la respectiva expansión de los factores.

A1. Control de calidad de la encuesta

Al momento de diligenciar la encuesta la persona necesitaba escribir su nombre, fecha, tiempo de inicio en que la contesta y cuando finaliza, información del grupo que lo acompañaba y por supuesto del supervisor responsable de la propiedad. En algunas ocasiones los entrevistadores volvieron al lugar de la encuesta por diferentes motivos, bien sea porque la persona no se encontraba, porque no se terminó la encuesta o porque hubo una abstinencia o rechazo en el momento de la encuesta.

B Datos de la vivienda y su entorno (para el primer hogar de la vivienda)

Datos del entorno

Esta sección específicamente describe los detalles sobre el lugar de la vivienda, si posee algún tipo de actividades económicas local, de industria o alguna otra actividad que presta algún tipo de servicio. También si tienen acceso a lugares con una iluminación. El tipo de pisos del hogar, si presenta grietas, humedad, hundimientos o si cuenta con cimientos, columnas o vigas.

Datos de la vivienda

Lo primero que se encuesta es el tipo de propiedad que se tiene, es decir si aquella es propia, en arriendo o demás, el tipo de construcción física, madera, ladrillos, zinc, piedra. El material predominante en mayor proporción para los pisos. También si se encuentra cerca de una zona industrial, aeropuertos, basureros, terminales y finalmente se tiene en cuenta cuáles son los problemas que se evidencian con mayor impacto, ruido, inseguridad, malos olores o invasión de andenes y calles.

C. Condiciones habitacionales del hogar (para todos los hogares de la vivienda)

La vivienda es propia y pagada completamente o parcial o si está en arriendo, información si se tuviera que pagar arriendo, la vivienda de quien esta a nombre, el año de la compra del mismo, si tienen intenciones de adquirir una vivienda y su financiación o las razones por las cuales no desea. Si ha recibido algún tipo de subsidio. Las divisiones frente a habitaciones, cómo se preparan los alimentos, y si se cuenta con una cocina, además si allí se utiliza algún tipo de energía o combustible para cocinar. Además de los recursos que se tiene, si se separan basuras, los bienes y al interior sobre prácticas de reducción de gastos.

D. Servicios públicos domiciliarios (para todos los hogares de la vivienda)

Aquí se presentan todos y cada uno de los servicios públicos, si tienen acceso a acueducto y alcantarillado, electricidad, gas natural y tecnologías de la información. Si se pagan por el hogar o entre varios hogares pertenecientes a la vivienda. Si ha habido algún inconveniente con la prestación del servicio. Para las TIC, si presentan computadores, el tipo de computador, televisión y el valor pagado por este servicio.

E. Composición del hogar y demografía (para todas las personas)

Es importante definir la relación de las personas con el jefe o jefa del hogar, así como el tipo de grupo étnico con el que se pueda representar, su estado civil actual, si vive con el o la cónyuge, si ha residido en Bogotá D.C. toda la vida, municipio de nacimiento, área del nacimiento, si ha vivido afuera de la ciudad, niveles de educación alcanzados por el padre y la madre.

F. Salud (para todas las personas del hogar)

Se encuentran inscritos a programas de salud, de seguridad social, el tipo de entidad si es EPS o salud subsidiada, los tipos de regímenes de salud, razones por las cuales no se encuentra inscrito, el tipo de subsidio, quién paga la EPS, si tiene un sueldo, cuánto le descuentan, si ha tenido inconvenientes con que lo reconozcan como afiliado, al igual si ha tenido algún tipo de enfermedad, que haya sido hospitalizada y la calidad del servicio que le hayan brindado. Si ha practicado algún tipo de deporte o parcialmente.

G. Cuidado de los niños y niñas menores de cinco años

Se desea saber si aquellos niños asisten a algún jardín, guardería u hogar comunitario, razón por la que no asiste, en el caso de no hacerlo, quién lo lleva usualmente al lugar donde asiste y de igual forma quién lo recoge al finalizar la jornada, duración del año escolar y valores estimados por pensiones y uniformes; si recibe comida allí, además quién ha permanecido el mayor tiempo con el niño durante la última semana. Si el niño ha tenido alguna enfermedad durante este tiempo.

H. Educación (para personas de cinco años y más)

Se pregunta si sabe leer y escribir, actualmente se encuentra estudiando, en preescolar, colegio o universidad; sin embargo también se quiere conocer si no se encuentra estudiando en el momento, explica los motivos por los cuales no se está haciendo, el nivel de educación alcanzado, en caso que haya terminado algún tipo de estudio, establecimiento, lugar donde se encuentra el establecimiento. Si es beneficiario de algún tipo de beca o subsidio por parte del gobierno. De igual forma el tiempo que transcurre en llegar al establecimiento. Si utiliza algún crédito para estudiar y las actividades que realizó el último mes en su tiempo libre.

I. Uso de las tecnologías de la información (para personas de diez años y más)

Este módulo tiene los componentes de interacción con la información, si la persona tiene un teléfono móvil con el acceso a internet y llamadas, de igual forma si tiene un computador, sitios en que se utiliza la mayoría del tiempo este computador. Si utiliza el internet la frecuencia o por qué no lo hace.

J. Participación en organizaciones y redes sociales (para personas de diez años y más)

Participación en reuniones sociales, o si pertenece a algún grupo o asociación. Dentro de la organización qué tipo de rol desempeña, el tiempo destinado, el tipo de organización, es decir su carácter social, o si no pertenece a algún tipo de grupo social, de igual forma a quién recurre al tener problemas económicos.

K. Fuerza de trabajo (para personas de diez años y más)

Involucra todas las posibles fuentes de ingresos que tienen las personas, es decir, si son empleados, independientes o reciben algún tipo de ingreso por una actividad económica. Cada una de estas personas pueden haber pagado por su trabajo, puede ser en efectivo y en especie, teniendo en cuenta el tiempo que se ha tomado el trabajo si la empresa ofrece transporte, alimentación u otro beneficio, o si no, cuánto tiempo lleva desempleado y si no ha pensado en hacer alguna actividad económica. Tenga en cuenta el lugar, o si usted tiene un negocio, finalmente, si esta persona tiene más personas a cargo.

En el ejercicio utilizamos las siguientes variables:

- **K1:** ¿En qué actividad ocupó la mayor parte del tiempo la semana pasada?
- **K2:** ¿Realizó la semana pasada alguna actividad paga por una hora o más?
- **K3:** ¿Tenía durante esa semana algún trabajo o negocio por el que recibe ingresos?
- **K4:** ¿Trabajó la semana pasada en un negocio por una hora o más, sin que le pagaran?
- **K15:** ¿Qué hace en este trabajo?
- **K28:** ¿En este trabajo se desempeña como?
- **K30:** ¿Cuánto ganó el mes pasado en este empleo (incluya propinas y comisiones y excluya viáticos y pagos en especie)?
- **K31:** ¿El mes pasado recibió alimentos como parte de pago por su trabajo?
- **K32:** ¿El mes pasado recibió vivienda como parte de pago por su trabajo?
- **K33:** ¿El mes pasado recibió otros ingresos en especie por su trabajo (electrodomésticos, ropa, productos diferentes a alimentos o bonos tipo Sodexho, etc.)?
- **K34:** ¿Normalmente utiliza transporte de la empresa para desplazarse a su trabajo (bus, automóvil), particular u oficial?
- **K35:** ¿El mes pasado recibió subsidio de alimentación en dinero?
- **K36:** ¿El mes pasado recibió auxilio de transporte en dinero?
- **K37:** ¿El mes pasado recibió subsidio familiar en dinero?
- **K38:** ¿El mes pasado recibió primas (técnica, de antigüedad, clima, orden público, etc.) en dinero?
- **K39A:** ¿Durante los últimos doce meses recibió prima de servicios?

- **K39B:** ¿Durante los últimos doce meses recibió prima de navidad?
- **K39C:** ¿Durante los últimos doce meses recibió prima de vacaciones?
- **K39D:** ¿Durante los últimos doce meses recibió bonificaciones?
- **K39E:** ¿Durante los últimos doce meses recibió pagos o indemnizaciones por accidentes de trabajo?
- **K40:** ¿Cuál fue la ganancia neta o los honorarios netos de ... en esa actividad, negocio, profesión o finca, el mes pasado?
- **K46:** ¿El mes pasado tuvo otros trabajos o negocios por los cuales recibió ingresos?
- **K51:** ¿El mes pasado recibió algún ingreso por concepto de pensión de jubilación, sustitución pensional, invalidez o vejez?
- **K52:** ¿El mes pasado recibió algún ingreso en dinero para el sostenimiento de hijos menores de dieciocho años (incluya pensión de alimentación y contribución de padres ausentes)?
- **K53:** ¿El mes pasado recibió algún ingreso por concepto de arriendos de casas, apartamentos, fincas de recreo, lotes, vehículos, maquinaria y equipo?
- **K54:** ¿Durante los últimos doce meses recibió primas por pensión de jubilación o por sustitución pensional?
- **K55:** ¿Durante los últimos doce meses recibió algún ingreso por concepto de ayudas en dinero provenientes de otros hogares o instituciones (padres, hijos, familiares, amigos)?
- **K57:** ¿Durante los últimos doce meses recibió dinero por otros conceptos (cesantías, intereses por cesantías, intereses por préstamos o CDT, rifas, etc.)?

L. Percepción sobre las condiciones de vida de la ciudad y el desempeño institucional (para el jefe de hogar o su cónyuge)

Se indaga sobre temas de victimización (en primera persona o con algún familiar), condiciones de vida, tipo de relación que se vivía en el hogar donde creció, ingresos, consideraciones que se tiene a sí mismo, si ha tenido problemas graves en la familia, percepción de mejora o empeoramiento de la ciudad. Condiciones de vida frente a la alimentación. Satisfacción con la vida en el barrio que se encuentra habitando.

M. Gastos

M1. Gastos en alimentos y bebidas no alcohólicas de los hogares (para todos los hogares de la vivienda)

Cuántas veces se acostumbran a hacer compras de comida y bebidas no alcohólicas en la vivienda, el dinero gastado para las mismas en los últimos siete días. Se adquirieron este tipo de comidas y bebidas sin tener que comprarlas y su forma de obtención.

M2. Otros gastos del hogar (para todos los hogares de la vivienda)

Otros gastos como gasolina, bebidas alcohólicas, correo electrónico y el precio que se pagó por cada uno de los servicios alternativos, como ropa, diversión y si se tuvo que pagar o no por estas actividades. Además del transporte, artículos del hogar y demás.

2.1. Variables para la estimación del índice de Gini

Cuando usamos la encuesta necesitamos diferente información, en este caso utilizaremos los módulos A, B y K. En el módulo A, tenemos el estrato, la localidad y la dirección por cada hogar; en el módulo B tendremos el factor de expansión, esto es la capacidad que tiene cada individuo seleccionado en una muestra probabilística para representar el universo en el cual se encuentra. Por lo que cuenta con la posibilidad de representar completamente el universo que hace parte del estudio, (DANE 2008).

El modulo K, contiene las variables específicamente acerca de los ingresos para cada uno de los hogares en Bogotá D.C., seguido de esto les mostraremos las variables que contienen este ingreso dividido por tipos de fuerza de trabajo.

Empleados

Estas personas responden las primeras cuatro preguntas, que son nuestras variables el modulo K, (K1, K2, K3, K4), que hace referencia a la pregunta de si se encuentra trabajando en este momento, seguido de esto nos lleva a la pregunta K15, donde la persona nos brinda el nombre del trabajo y en la K28 nos dice el tipo de trabajo que realiza en las opciones 1, 2, 3 y 11 su trabajo. Finalmente se observan todos los ingresos que ha tenido como subsidios o algún otro tipo de soporte económico. En K30 la persona nos informa acerca de los ingresos mensualmente. Es importante dividir en doce las variables que se encuentran expresadas en años, porque el principal cálculo del ingreso está expresado en años. A continuación mostramos las variables K30, K31, K32, K33, K34, K35, K36, K37, K38, K39A, K39B, K39C, K39D, K39E, K46, K51, K52, K53, K54, K55, K57, K30, K31, K32, K33, K34, K35, K36, K37, K38, K39A, K39B, K39C, K39D, K39E, K46, K51, K52, K53, K54, K55, K57 para el cálculo de los ingresos en cada hogar.

Independientes

Para los independientes usamos un filtro similar con las respuestas de las primeras cuatro variables (K1, K2, K3, K4), pero en K28 respondieron la opción 4, 5, 6, y

7, acerca de su trabajo independiente. En adición mostraremos las variables que nos describen esta actividad K40, K46, K51, K52, K53, K54, K55, K57. Ocurre la misma situación, sobre la división de las variables en 12 para tener su valor mensual y no anual, K54, K55, K57.

Trabajadores sin pago

Los trabajadores sin pago son seleccionados en las primeras cuatro preguntas (K1, K2, K3, K4) y que respondieron en la variable K28 las opciones 8, 9 y 10, de igual forma se seleccionarán K46, K51, K52, K53, K54, K55, K57. Después de esto es necesario igualmente dividir en 12, K54, K55, K57.

Desempleados

Las personas que pertenecen a este grupo son definidas por las opciones 2, 3, 4, 5, y 6 en la pregunta K1. Aquí justamente también se debe dividir en 12 la variable K57. Necesitamos las variables K46, K51 y K52, para diferenciarlos de los demás tipos de trabajadores en la encuesta.

Nota: la población fue dividida en cuatro grupos mutuamente excluyente por grupos en los cuales el rol de cada persona es garantizado por la Encuesta Multipropósito para Bogotá D.C., 2011.

3. Manipulación de los datos en R

En esta sección el lector podrá visualizar el código utilizado en el cálculo del índice de Gini en R, para cada una de las localidades urbanas, a excepción de Sumapaz, que es rural; recordemos que este se basa en la utilización de la Encuesta Multipropósito para Bogotá D.C., 2011.

El primer paso para el cálculo es descargar las siguientes librerías con sus respectivos paquetes: librería `foreign` es utilizada para leer los datos que provienen del software SPSS y otras extensiones, librería `reldist` con la cual se calcula el índice de Gini, librería `gtools` la utilizamos para la función `Smartbind`, con la cual combinamos las columnas, la librería `Hmisc` se utiliza para obtener estadísticas básicas por variables, librería `plyr`, suma los pesos de las variables dadas y la librería `sqldf` para la lectura de archivos SQL en R, usados a conveniencia.

```
library(foreign)
library(reldist)
library(gtools)
library(Hmisc)
library(plyr)
library(sqldf)
```

Leemos los datos con la función SPSS de la librería `foreign`:

```
setwd("../..")
A=read.spss("capa.sav",use.value.labels=T,to.data.frame=T)
B=read.spss("capb.sav",use.value.labels=F,to.data.frame=T)
k=read.spss("capk.sav",use.value.labels=F,to.data.frame=T)
```

Aquí utilizamos la organización del modulo K para cada uno de los hogares con el número de su directorio, para que sea más sencillo identificar las localidades de cada hogar y substraer de allí cada persona:

```
k$DIRECTORIO = as.character(k$DIRECTORIO)
k$DIRECTORIO_PER = as.character(k$DIRECTORIO_PER)
k$DIRECTORIO_HOG = NULL
k$DIRECTORIO_HOG = substr(k$DIRECTORIO_PER,1,8)
```

La función `convert` para convertir variables numéricas del modulo K, remueve los espacios encontrados después de los datos al convertirse en caracteres y números:

```
convert = function(x){
  x = as.character(x)
  x = gsub(" ", "", x)
  x = as.numeric(x)
}

k$DIRECTORIO = as.character(k$DIRECTORIO)
k$DIRECTORIO_PER = as.character(k$DIRECTORIO_PER)
k$DIRECTORIO_HOG = as.character(k$DIRECTORIO_HOG)
table(duplicated(k$DIRECTORIO_HOG))
table(duplicated(k$DIRECTORIO_PER))
```

Definimos K para que tome los valores de las 112 variables:

```
K = k
for(j in 1:112 ){
  K[,j]=convert(K[,j])
}
```

Compacta la estructura interna en nuestra nueva base K y cuenta el número de características para la variable:

```
str(K)
table(nchar(K$DIRECTORIO_PER))
```

El siguiente paso es realizar el filtro en cada categoría, empleados, independientes, trabajadores sin pago y desempleados con las variables nombradas en la sección 2.1.:

Empleados: el filtro que define los empleados se encuentra a través de las preguntas K1 (opción 1), K2 (opción 1), K3 (opción 1), K4 (opción 1), en las opciones 1, 2, 3, 11 para la variable K28, creando una nueva variable llamada *Asal* y estas variables son convertidas mensualmente dado que se recogieron anualmente:

```
attach(K)
filtro_asal<- (K1==1 | K2==1 | K3==1 | K4==1) & K28 %in% c(1,2,3,11)
detach(K)
Asal =
subset(K,filtro_asal,
  select=c(DIRECTORIO_HOG,DIRECTORIO_PER,K1,
    K2,K3,K4,K28,K30_VALOR_MES,K31_VALMES_EST,K32_VALMES_EST,
    K33_VALMES_EST,K34_VALMES_EST,K35_VALOR,K36_VALOR,K37_VALOR,
    K38_VALOR,K39A_VALOR,K39B_VALOR,K39C_VALOR,K39D_VALOR,
    K39E_VALOR,K46_VALOR,K51_VALOR,K52_VALOR,K53_VALOR,K54_VALOR,
    K55_VALOR,K57_VALOR))

Asal$K39A_VALOR = Asal$K39A_VALOR/12
Asal$K39B_VALOR = Asal$K39B_VALOR/12
Asal$K39C_VALOR = Asal$K39C_VALOR/12
Asal$K39D_VALOR = Asal$K39D_VALOR/12
Asal$K39E_VALOR = Asal$K39E_VALOR/12
Asal$K54_VALOR = Asal$K54_VALOR/12
Asal$K55_VALOR = Asal$K55_VALOR/12
Asal$K57_VALOR = Asal$K57_VALOR/12
```

Independientes: el filtro que define la variable se encuentra en las preguntas K1 (opción 1), K2 (opción 1), K3 (opción 1), K4 (opción 1), en las opciones 4, 5, 6 y 7 de la variable K28, creando una nueva variable llamada *Indep*, de igual forma se convirtieron en meses ya que proceden en años:

```
attach(K)
filtro_indep = (K1==1 | K2==1 | K3==1 | K4==1) & K28 %in% c(4,5,6,7)
detach(K)
Indep = subset(K,filtro_indep,
  select=c(DIRECTORIO_HOG,DIRECTORIO_PER,K1,
    K2,K3,K4,K28,K40_GANACIA,K46_VALOR,K51_VALOR,
    K52_VALOR,K53_VALOR,K54_VALOR,K55_VALOR,K57_VALOR))

Indep$K54_VALOR = Indep$K54_VALOR/12
Indep$K55_VALOR = Indep$K55_VALOR/12
Indep$K57_VALOR = Indep$K57_VALOR/12
```

Trabajadores sin pago: el filtro esta definido para los trabajadores sin remuneración en las preguntas K1 (opción 1), K2 (opción 1), K3 (opción 1), K4 (opción 1), en las opciones 8 y 10 de la variable K28, creando así la nueva variable `TrabSinrem` y convirtiendo las variables anuales en mensuales:

```
attach(K)
filtro_trabsinrem <- (K1==1 | K2==1 | K3==1 | K4==1) & K28 %in% 8:10
detach(K)
TrabSinrem = subset(K,filtro_trabsinrem,
  select=c(DIRECTORIO_HOG,DIRECTORIO_PER,K1,
    K2,K3,K4,K28,K46_VALOR,K51_VALOR,
    K52_VALOR,K53_VALOR,K54_VALOR,
    K55_VALOR,K57_VALOR))

TrabSinrem$K54_VALOR = TrabSinrem$K54_VALOR/12
TrabSinrem$K55_VALOR = TrabSinrem$K55_VALOR/12
TrabSinrem$K57_VALOR = TrabSinrem$K57_VALOR/12
```

Desempleados: definiendo esta variable, bajo el filtro de la pregunta K1 en las opciones 2 y 6, creando la nueva variable llamada `TrabDesoc`, estas variables se convierten en mensuales ya que vienen en años:

```
attach(K)
filtro_desoc <- K1 %in% 2:6
detach(K)
TrabDesoc = subset(K,filtro_desoc,
  select=c(DIRECTORIO_HOG,DIRECTORIO_PER,
    K46_VALOR,K51_VALOR,
    K52_VALOR,K57_VALOR))

TrabDesoc$K57_VALOR = TrabDesoc$K57_VALOR/12
```

Indicando que se toman valores para cada una de las variables requeridas para el índice de Gini:

```
Asal$indica_ocupa = "Asalariado"
Indep$indica_ocupa = "Independiente"
TrabSinrem$indica_ocupa = "Trabajador sin remuneracion"
TrabDesoc$indica_ocupa = "Desocupado o inactivo"
```

Aquí utilizamos la librería `gtools` con su función `smartbind`, con el propósito de agregar las últimas variables en su descripción para los ingresos y substraer los primeros seis dígitos del directorio de los hogares:

```
PersonasTrab = smartbind(Asal,Indep,TrabSinrem,TrabDesoc)
PersonasTrab$DIRECTORIO = substr(PersonasTrab$DIRECTORIO_HOG,1,6)
```

rowSums, calcula las sumas de las columnas a través de las filas de una matriz numérica de un objeto para cada nivel de grupos de variables, en este caso para cada persona, tomando los nombres de las filas o de otras versiones de la base de datos:

```
PersonasTrab= merge(PersonasTrab,A,by="DIRECTORIO",all.x=T)
PersonasTrab$Ingreso = rowSums (PersonasTrab[,c(4:29,31)],na.rm=T)
```

Esta parte es muy importante porque se utiliza el factor de expansión que es usado con los hogares:

```
PersonasTrab = merge (PersonasTrab,B [c("DIRECTORIO", "FEX_C")],
  by="DIRECTORIO", all.x=T)
```

Combinamos los resultados para cada hogar sumando sus ingresos y pesos con la librería plyr:

```
weighted.sum = function(x,w,na.rm=T){
  sum(x*w,na.rm=na.rm)
}
```

Creamos la suma de las entradas de los ingresos para el hogar y es combinado con los factores de expansión obtenidos con la librería sqldf:

```
BaseIngr_Hog = sqldf("select DIRECTORIO_HOG, sum(Ingreso)
  as 'IngresoHog' from PersonasTrab group by DIRECTORIO_HOG")
BaseIngr_Hog$DIRECTORIO = substr(BaseIngr_Hog$DIRECTORIO_HOG,1,6)
BaseIngr_Hog = merge(BaseIngr_Hog,A,by="DIRECTORIO",all.x=T)
BaseIngr_Hog = merge(BaseIngr_Hog,B[c("DIRECTORIO", "FEX_C")],
  by="DIRECTORIO",all.x=T)
```

Se toman los ingresos del hogar traídos por el directorio:

```
BaseIngr_Hog$DIRECTORIO = substr(BaseIngr_Hog$DIRECTORIO_HOG,1,6)
describe(BaseIngr_Hog)
```

Se remueven los espacios contenidos en la base de datos:

```
BaseIngr_Hog$LOCALIDAD = as.character(BaseIngr_Hog$LOCALIDAD)
BaseIngr_Hog$LOCALIDAD = gsub(" ", "", BaseIngr_Hog$LOCALIDAD)
```

Creamos una nueva columna con el nombre de cada localidad urbana de Bogotá D.C. para el trabajo, recordando que solo es en las diecinueve localidades, excluyendo a Sumapaz:

```
nombreslocal = read.csv("nombreslocal.csv", sep=";")
nombreslocal$LOCALIDAD = as.character(nombreslocal$LOCALIDAD)
BaseIngr_Hog = merge(BaseIngr_Hog, nombreslocal, by="LOCALIDAD")
BaseIngr_Hog$LOCALIDAD = ifelse(nchar(BaseIngr_Hog$LOCALIDAD)==1,
                                paste0("0", BaseIngr_Hog$LOCALIDAD),
                                BaseIngr_Hog$LOCALIDAD)
```

Comenzamos con el cálculo del índice de Gini con la librería `reldist`:

```
GINITOTAL = gini(x = BaseIngr_Hog$IngresoHog, weights=BaseIngr_Hog$FEX_C)
```

Nombramos las localidades y agregamos su número para cada una:

```
USAQUEN = subset(BaseIngr_Hog, LOCALIDAD=="01")
CHAPINERO = subset(BaseIngr_Hog, LOCALIDAD=="02")
SANTAFE = subset(BaseIngr_Hog, LOCALIDAD=="03")
SANCRISTOBAL = subset(BaseIngr_Hog, LOCALIDAD=="04")
USME = subset(BaseIngr_Hog, LOCALIDAD=="05")
TUNJUELITO = subset(BaseIngr_Hog, LOCALIDAD=="06")
BOSA = subset(BaseIngr_Hog, LOCALIDAD=="07")
CIU_KENNEDY = subset(BaseIngr_Hog, LOCALIDAD=="08")
FONTIBON = subset(BaseIngr_Hog, LOCALIDAD=="09")
ENGATIVA = subset(BaseIngr_Hog, LOCALIDAD=="10")
SUBA = subset(BaseIngr_Hog, LOCALIDAD=="11")
BARR_UNIDOS = subset(BaseIngr_Hog, LOCALIDAD=="12")
TEUSAQUILLO = subset(BaseIngr_Hog, LOCALIDAD=="13")
LOSMARTIRES = subset(BaseIngr_Hog, LOCALIDAD=="14")
ANT_NARINO = subset(BaseIngr_Hog, LOCALIDAD=="15")
PTE_ARANDA = subset(BaseIngr_Hog, LOCALIDAD=="16")
CANDELARIA = subset(BaseIngr_Hog, LOCALIDAD=="17")
RAF_URIBE = subset(BaseIngr_Hog, LOCALIDAD=="18")
CDAD_BOLIVAR = subset(BaseIngr_Hog, LOCALIDAD=="19")
```

Agregamos el factor de expansión proveniente de los ingresos en cada localidad:


```

GINITOTAL_USAQUEN = gini(x = USAQUEN$IngresoHog,
                        weights=USAQUEN$FEX_C)
GINITOTAL_CHAPINERO = gini(x = CHAPINERO$IngresoHog,
                          weights=CHAPINERO$FEX_C)
GINITOTAL_SANTAFE = gini(x = SANTAFE$IngresoHog,
                        weights=SANTAFE$FEX_C)
GINITOTAL_SANCRISTOBAL = gini(x = SANCRISTOBAL$IngresoHog,
                              weights=SANCRISTOBAL$FEX_C)
GINITOTAL_USME = gini(x = USME$IngresoHog,
                     weights=USME$FEX_C)
GINITOTAL_TUNJUELITO = gini(x = TUNJUELITO$IngresoHog,
                            weights=TUNJUELITO$FEX_C)
GINITOTAL_BOSA = gini(x = BOSA$IngresoHog,
                    weights=BOSA$FEX_C)
GINITOTAL_CIU_KENNEDY = gini(x = CIU_KENNEDY$IngresoHog,
                             weights=CIU_KENNEDY$FEX_C)
GINITOTAL_FONTIBON = gini(x = FONTIBON$IngresoHog,
                          weights=FONTIBON$FEX_C)
GINITOTAL_ENGATIVA = gini(x = ENGATIVA$IngresoHog,
                          weights=ENGATIVA$FEX_C)
GINITOTAL_SUBA = gini(x = SUBA$IngresoHog,
                    weights=SUBA$FEX_C)
GINITOTAL_BARR_UNIDOS = gini(x = BARR_UNIDOS$IngresoHog,
                             weights=BARR_UNIDOS$FEX_C)
GINITOTAL_TEUSAQUILLO = gini(x = TEUSAQUILLO$IngresoHog,
                             weights=TEUSAQUILLO$FEX_C)
GINITOTAL_LOSMARTIRES = gini(x = LOSMARTIRES$IngresoHog,
                              weights=LOSMARTIRES$FEX_C)
GINITOTAL_ANT_NARINO = gini(x = ANT_NARINO$IngresoHog,
                             weights=ANT_NARINO$FEX_C)
GINITOTAL_PTE_ARANDA = gini(x = PTE_ARANDA$IngresoHog,
                             weights=PTE_ARANDA$FEX_C)
GINITOTAL_CANDELARIA = gini(x = CANDELARIA$IngresoHog,
                             weights=CANDELARIA$FEX_C)
GINITOTAL_RAF_URIBE = gini(x = RAF_URIBE$IngresoHog,
                           weights=RAF_URIBE$FEX_C)
GINITOTAL_CDAD_BOLIVAR = gini(x = CDAD_BOLIVAR$IngresoHog,
                              weights=CDAD_BOLIVAR$FEX_C)

```

Creamos el nuevo vector con todas las localidades:

```

GINIS = c(GINITOTAL, GINITOTAL_USAQUEN, GINITOTAL_CHAPINERO,
          GINITOTAL_SANTAFE, GINITOTAL_SANCRISTOBAL, GINITOTAL_USME,
          GINITOTAL_TUNJUELITO, GINITOTAL_BOSA, GINITOTAL_CIU_KENNEDY,
          GINITOTAL_FONTIBON, GINITOTAL_ENGATIVA, GINITOTAL_SUBA,

```

```

,GINITOTAL_BARR_UNIDOS,GINITOTAL_TEUSAQUILLO,
,GINITOTAL_LOSMARTIRES,GINITOTAL_ANT_NARINO,
GINITOTAL_PTE_ARANDA,GINITOTAL_CANDELARIA,)
GINITOTAL_RAF_URIBE,GINITOTAL_CDAD_BOLIVAR)
BASE_GINI = data.frame(LOCALIDAD = 0:19 ,
  NOMBRES = c("TOTAL", "USAQUEN", "CHAPINERO", "SANTAFE",
  "SANCRISTOBAL", "USME", "TUNJUELITO", "BOSA",
  "KENNEDY", "FONTIBON", "ENGATIVA", "SUBA",
  "BARR_UNIDOS", "TEUSAQUILLO", "LOSMARTIRES",
  "ANT_NARINO", "PTE_ARANDA", "CANDELARIA",
  "RAF_URIBE", "CDAD_BOLIVAR"),GINI = GINIS)

BASE_GINI$LOCALIDAD = ifelse(nchar(BASE_GINI$LOCALIDAD)==1,
  paste0("0",BASE_GINI$LOCALIDAD),BASE_GINI$LOCALIDAD)

BASE_GINI=BASE_GINI[-1,]

```

Finalmente obtenemos el índice de Gini para cada localidad con su factor de expansión por ingresos y su ubicación en el orden, con la siguiente línea, de la cual se obtiene la tabla por cada localidad:

```
BASE_GINI
```

4. Discusión

De esta forma la diferencia entre la comparación y la desigualdad de la información, en economía, tiene una respuesta que expresa estas características es el índice de Gini. Madden (2000) muestra la distribución que tiene el índice el cual toma valores entre 0 y 1, mostrándonos la diferencia entre una completa distribución de los ingresos y una perfecta desigualdad de los mismos para los hogares en Bogotá D.C..

La complejidad en la interpretación se encuentra en la forma de analizar la razón por la cual el índice de Gini presenta una gran sensibilidad frente a cambios diversos en la parte media de la distribución y en los extremos finales. La causa de este suceso tiene que ver con el hecho de que la población con el promedio de ingresos tiene una gran proporción de ingresos en unas pequeñas minorías de las personas.

La principal razón para calcular el índice de Gini para las localidades consiste en entender como es la distribución de los ingresos en ellas, para eso utilizamos la calidad de vida y la diferencia que se tienen entre cada uno, puesto que nos muestran características particulares relacionados con el ingreso. Se encuentran diferentes categorías donde se ubica la población de Bogotá D.C., como lo son los emplea-

Tabla 1: *Estimación del índice de Gini para cada localidad. Fuente: elaboración propia.*

Número	Localidad	Gini
1	Usaquén	0.5619379
2	Chapinero	0.5729119
3	Santa Fe	0.5633526
4	San Cristobal	0.4379722
5	Usme	0.4068284
6	Tunjuelito	0.4622931
7	Bosa	0.4201847
8	Kenndy	0.4459290
9	Fontibón	0.5245685
10	Engativa	0.4486164
11	Suba	0.5179652
12	Barrios Unidos	0.5276923
13	Teusaquillo	0.5064627
14	Los Mártires	0.5341255
15	Antonio Nariño	0.5150606
16	Pte. Aranda	0.4523843
17	Candelaria	0.6055275
18	Rafael Uribe	0.4941516
19	Ciudad Bolívar	0.4222946

dos, independientes, trabajadores sin pago y los desempleados son segmentos en los cuales se obtuvo el Gini para cada localidad de la ciudad.

Para entender cómo se distribuye la capital de Colombia, el índice de Gini, provee una clara indicación de la inclinación que tienen los ingresos en las diecinueve localidades, recordemos que nosotros no incluimos la localidad No. 20 que es Sumapaz, ya que su ubicación se encuentra fuera de nuestra área de interés por ser rural, en Lewis (2012) dice que el índice de Gini es una medida apropiada para usarse cuando queremos medir los efectos de las políticas públicas, frente a la distribución de los ingresos.

Cuando tenemos la posibilidad de observar los valores que está tomando el índice de Gini para cada localidad, entendemos las situaciones tan diversas en las que se encuentra Bogotá D.C., porque de acuerdo con Litchfield (1999) la desigualdad es también un concepto que tiene que ver con el bienestar. Esta información sobre la distribución, indica que los hogares cuentan con características específicas sobre el tipo de vida y la diferencia socioeconómica en cada hogar que se encuentra en cada localidad.

En este orden Usaquén tiene un valor de 0.5632 del índice de Gini, y expresa un valor de desigualdad bastante amplio ya que se encuentra por encima de la mitad de los valores que puede tomar el índice, evidenciando que hay un número considerable de personas que están congregando el ingreso en mayor proporción

que el resto de los hogares que pertenecen a esta localidad. Es la primera localidad en la que se debería generar un adecuado sistema en la ciudad, como transporte, acueducto o espacio público, etc., esta localidad tiene una agrupación de personas con una formación de calidad intelectual en lo que tiene que ver con arte, música y una alta educación; además son saludables, esta última característica incluye deportes y demás clases de ejercicio. Tiene un gran porcentaje de personas adultas de acuerdo con Secretaria Distrital de Planeación (2009).

En particular para algunas localidades el índice de Gini es similar: Chapinero 0.5729119, Santa Fe 0.5628, Fontibón 0.5245, Antonio Nariño 0.5169, Suba 0.5210, Barrios Unidos 0.5275, Teusaquillo 0.5057 y Los Mártires 0.5358. Es posible que estas localidades tengan algunos hogares que presentan una concentración del dinero similar, siendo altos los ingresos; por ejemplo: Chapinero tienen un porcentaje importante de lugares que prestan algún tipo de servicio y de comercio, su población se encuentra al rededor de 166000 personas, es considerada como un espacio cultural de Bogotá D.C., posee una gran variedad de universidades, donde el flujo de dinero es bastante alto, sus barrios son bastantes variados, teniendo sectores exclusivos, siendo esto una causa de que su índice de Gini sea alto y cercano a 1, como se presenta en la localidad de Santa Fe.

Fontibón, Teusaquillo, Suba, Los Mártires, Antonio Nariño y Barrios Unidos tienden a presentar comportamientos similares, en el caso de los ingresos, porque todas las localidades nombradas tienen unas grandes contribuciones frente a servicios y comercio. Con respecto a la población, también existe una alta proporción, siendo localidades que albergan gran cantidad de ciudadanos, una característica que vale la pena resaltar es la variedad que presenta frente a los estratos, ya que congrega altos y bajos. Es importante tener en cuenta que estos estratos altos son los que congregan la mayor cantidad de ingresos en estas localidades.

Las otras localidades con un bajo índice de Gini, indican una congregación más equilibrada frente a la distribución de los ingresos, estratos similares, formas de trabajo parecidas, y con ubicaciones al sur de la ciudad, en su mayoría. Sin embargo, tenemos el caso que es la excepción con el índice de Gini más alto entre todas las localidades de Bogotá D.C. (la Candelaria), y es debido a que es una zona que alberga gran capacidad hotelera, donde unas pocas personas tienden a ser los dueños de estos lugares, del comercio como restaurantes, bares, museos y además zonas residenciales de gran magnitud, sin dejar de lado que se encuentran las universidades más costosas en su zona. Lo anterior pone de manifiesto una ley de atracción frente a la concentración de dinero, en comparación con las demás localidades.

Finalmente podemos pensar que el índice de Gini, es una gran medida de desigualdad que nos indica el comportamiento y aplicación de aquellas políticas públicas impuestas por personeros y alcaldes locales, de la mano con el Alcalde Mayor de la ciudad de Bogotá D.C.. En donde encontramos una gran marca de desigualdad, de concentración de ingresos en unas pocas personas que estarían teniendo el control de la mayoría de recursos económicos, dejando así una pequeña parte de este a hogares con desventajas para una competencia legal.

Recibido: 23 de marzo de 2015

Aceptado: 13 de mayo de 2015

Referencias

- Chotikapanich, D. (2010), *Modeling Income Distributions and Lorenz Curves*, Springer.
- DANE (2008), 'Especificaciones de coeficiente y varianza encuesta de consumo cultural -ecc-', *Departamento Administrativo Nacional de Estadística* .
- Handcock, M. S. & Morris, M. (1999), *Relative Distribution Methods in the Social Sciences (Statistics for Social and Behavioral Sciences)*, Springer.
- Lewis, M. (2012), *Applied Statistics for Economists*, Routledge.
- Litchfield, J. (1999), 'Inequality: Methods and tools', *World Bank's Web Site* .
- Madden, J. F. (2000), *Changes in Income Inequality Within U.S. Metropolitan Areas*, W.E. Upjohn Institute.
- Mark, S. (1999), Relative distribution methods in the social sciences, *in* 'Statistics for Social Science and Public Policy', Springer-Verlag.
- Secretaria Distrital de Planeación (2009), 'Conociendo la localidad de usaquen: diagnóstico de los aspectos físicos, demográficos y socioeconómicos'.