

CÁLCULO MUESTRAL ESTRATIFICADO CON AFIJACIÓN PROPORCIONAL AL TAMAÑO PARA EL ANÁLISIS DE CONSUMO, DESPLAZAMIENTOS E IDENTIDAD LOCAL EN EL DISTRITO METROPOLITANO DE QUITO

STRATIFIED SAMPLING CALCULATION METHOD WITH PROPORTIONAL ALLOCATION TO THE SIZE FOR THE ANALYSIS OF CONSUMPTION, COMMUTATION AND LOCAL IDENTITY WITHIN THE METROPOLITAN DISTRICT OF QUITO

Puente Fernando / fpuente@uce.edu.ec
Hurtado Diego / dehurtado@uce.edu.ec
Morillo Jenny / jmmorillo@uce.edu.ec
Díaz Mayra / mayacsb@hotmail.com
Paz Gabriela / paolitapaz08@gmail.com

RESUMEN. Al no tener a la fecha del estudio un número exacto de habitantes en el Distrito Metropolitano de Quito (DMQ), contemplada desde los 10 hasta los 65 años de edad, se delimitó el área del estudio tomando como base el Censo Poblacional INEC del año 2010, mediante la selección de los polígonos censales en suelo de clasificación urbana para obtener una aproximación del tamaño de población de 1655073.

Se realizó mediante Muestreo Estratificado con afijación proporcional al tamaño de subpolígonos censales: celdas en todo el Distrito, de 400*400 metros cuadrados considerada como el área caminable, los cuales fueron analizados como estratos resultando un total de 2824 celdas. Para la división estratificada del Distrito Metropolitano de Quito se consideró la densidad poblacional de cada celda para determinar el número de encuestas, excluyendo las áreas verdes, parques, quebradas, estadios y centros educativos de gran extensión por no tener significancia poblacional. Obteniendo un total de 10503 encuestas georeferenciadas por el número identificador de lote.

Se determinó el peso de los estratos dentro del universo poblacional del DMQ en suelo urbano, estableciendo el número de encuestas que se debe aplicar dentro de cada celda para efectos del análisis de consumo, desplazamientos e identidad local. En particular se destacan los siguientes resultados: 25 encuestas a ser aplicadas en 27 celdas y 33 encuestas en 2 celdas, evidenciando estos como los espacios donde se está la mayor densidad local dentro del DMQ.

PALABRAS CLAVES: Censo, Centralidad Local, División Estratificada, Celdas, Muestra, Población, Polígonos Censales

ABSTRACT. Due to the lack of a number of inhabitants at the time of the study in the Metropolitan District of Quito (DMQ), considering ages from them 10 to 65, the study area was delimited taking as base the population Census INEC from the year 2010, through the selection of the census tracts in urban land classification to obtain an approximation of the population size, about 1655073.

The study was held by stratified sampling with proportional affixation to the size, taking into account the Census polygons sub-forming cells throughout the district, a grid composed of 400*400 square meters cells, considered as the walking area, which were analyzed as strata. The total number of cells was 2824.

For the stratified division of the Metropolitan District of Quito it was considered the population density of each cell to determine the number of surveys, excluding green areas, parks, ravines, stadiums and large-scale educational facilities, because they do not have population significance. A total of 10503 surveys were georeferenced by the lot identification number as a result.

It was determined the weight of the strata within the population universe of the DMQ in urban land, setting the number of surveys that apply within each cell for purposes of the analysis of consumption, commuting and local identity. Particularly it is highlighted the following results: a total of 25 surveys to be applied in 27 cells and 33 surveys in 2 cells, which evidence these spaces as the ones with the highest local density within the DMQ.

KEY WORDS. Census, Local Centrality, Stratified Division, Cells, Sample, Population, Census tracts.

INTRODUCCIÓN

Unas de las obras más destacadas en cuanto al estudio del muestreo por Juan Argibay “muestra” la forma de seleccionar a los sujetos y el tamaño de la muestra, tienen una importancia fundamental al momento de poder determinar en qué medida esa muestra es representativa de la población a la cual refiere. Que el muestreo sea probabilístico, es un tema sumamente importante en el proceso de garantizar una muestra no sesgada, lo cual, en principio, se vincula con la validez externa de la investigación. (Argibay, 2009)

Trabajar con muestras puede permitir realizar un análisis inferencial más exhaustivo y no un estudio meramente exploratorio, cave recalcar que en la selección de una muestra es importante considerar que esta sea lo suficientemente representativa para luego hacer inferencias a la población. Para obtener unos resultados representativos, es necesario que el tipo de muestreo que se utilice en una investigación, reúna los requisitos del análisis inferencial.

El fin de realizar un diseño muestral es hacer inferencias siendo válidas para obtener conclusiones extendidas a la población en base a la muestra, permitiendo reducir costos y tiempo. Para que el muestreo sea representativo se considera los siguientes criterios: Orientación hacia la meta, la medibilidad, la practicidad y la economía.

Determinar el número aproximado de habitantes dentro de la parte urbana del DMQ, considerando la población contemplada desde los 10 años en adelante para obtener una base de datos acorde a los requerimientos de nuestra investigación donde se identifica las diversas centralidades locales dentro del DMQ.

La estimación conjunta de totales poblacionales para distintas variables de interés en encuestas multi-propósito que utilizan diseños de muestreo estratificados. En particular, se proponen distintos métodos de estimación insesgada cuando el contexto del problema induce una población con una estructura jerárquica. (Argibay, 2009).

El cálculo de la población mediante los polígonos censales y la muestra en base a un diseño muestral Estratificado, eficiente y significativo para la población urbana del Distrito Metropolitano de Quito (DMQ) basados en el censo del INEC 2010 tomando como macro a la provincia de Pichincha y micro el área urbana y rural del DMQ mediante la aplicación del software arcGIS que permite analizar, organizar, recopilar y distribuir información geográfica.

El Modelo General para la Medición de Centralidad de Identidad Local. Mediante la noción de red que involucra no solamente la identificación de los nodos de la red, sino también los vínculos. Esta identifi-

cación de la red requiere necesariamente la asociación al concepto de Topología, y de lo cual se puede encontrar una referencia pronta en el trabajo de Rómulo Krafta en su análisis sobre Porto Alegre. En éste estudio, Krafta interpreta el territorio a través de una malla topológica que contiene entidades (nodos) vinculadas entre ellas, cuya relación entre ellas y sus atributos definen el grado de centralidad de una determinación territorial. (Bravo, 2001), siendo el nodo un espacio que forma parte de un todo y está entrelazada, dentro de la investigación planteada se realiza subdivisiones a las cuales llamaremos celdas cuya finalidad es tener la menor área de estudio e identificar las centralidades locales dentro del DMQ para realizar el análisis de consumo, desplazamientos e identidad local.

Calcular la muestra mediante un muestreo probabilístico estratificado, considerando el nivel de significancia, el error, la máxima variabilidad, la tasa de no respuesta. Para el cálculo de los estratos se consideró las distancias caminables de 400m por 400m y para el efecto del levantamiento de la información no se consideró las áreas verdes, Parques, Quebradas, Estadios, Centros de estudio.

La inclusión del espacio en los modelos matemáticos se ha acelerado con el desarrollo de los sistemas de información geográfica, lo que se ha traducido en un considerable avance metodológico, especialmente de la estadística espacial. La disposición espacial de las unidades, lo que desde la geografía urbana o la sociología denominaríamos barrios, puede ser analizada mediante su inclusión en las bases de datos georeferenciados. (Argibay, 2009)

Dentro de la encuesta se encuentra el número de predio para determinar el lugar exacto de residencia de la persona encuestada, consumo local, actividades diarias y su relación el barrio, uso del espacio público el mismo que para su aplicación se requiere de un diseño muestral.

El diseño muestral está basado en una investigación exploratoria debido a que no tenemos datos precisos de la población de estudio, la misma que nos permite obtener nuevos datos y elementos que puedan familiarizarnos con el número aproximado de la población del DMQ y poder sacar la muestra aproximada de habitantes para hacer el levantamiento de la información mediante encuestas las mismas que servirán para determinar las centralidades y la identidad local.

METODOLOGÍA

Calcular la muestra para determinar la identificación barrial, la centralidad y los lugares de mayor concentración de la gente de acuerdo a la actividad principal sea estudios, trabajo u otros.

Define la población como la totalidad de valores posi-

bles (mediciones o conteos) de una característica particular de un grupo especificado de objetos. Tal grupo especificado de objetos de grupo se llama universo.

La población puede ser finita o infinita, donde consideramos la primera como aquella que se puede determinar el número de individuos que la conforma, mientras en la infinita la ejemplifica como los resultados posibles de lanzar una moneda (cara o cruz). (Spiegel & Stephens, 2002, pág. 1)

Hay dos grandes grupos de muestreo: Probabilísticos, todos los individuos o elementos de la población tienen la misma probabilidad de ser incluidos en la muestra extraída, asegurándonos la representatividad de la misma y No Probabilísticos, por su parte, los elementos de la muestra se seleccionan siguiendo criterios determinados siempre procurando la representatividad de la muestra. (Kish, 1972, pág. 35)

En muchos de los casos, la realización de un censo no es posible por ser muy costoso, muy extenso. En tales oportunidades se debe practicar un análisis muestral. La muestra es una parte seleccionada de la población que deberá ser representativa, es decir, debe reflejar adecuadamente las características que deseamos analizar en el conjunto en estudio. (Kish, 1972, pág. 100)

La muestra debe ser representativa a esa población, con el fin de analizar mediante inferencia acerca de la misma. A la parte de la estadística que se ocupa de las condiciones bajo las cuales tales inferencias son válidas se llama estadística inductiva o inferencial. Debido a que tales inferencias no llegan a ser precisas por completo, para obtener conclusiones se utilizan probabilidades. La muestra es una parte de la población, seleccionada de acuerdo a una regla o plan. (Ostle, 1974, pág. 64)

Para obtener el total de población del distrito metropolitano de Quito se utilizó como base al censo del INEC 2010, Dirección de Información Cartográfica Estadística -DICE- a nivel de zonas censales del Distrito Metropolitano de Quito, que contiene la base de datos actualizada de los códigos DPA-División Política Administrativa, adicionalmente, se solicitó la entrega de los códigos DPA bajo el mismo formato anterior pero que incluyan las parroquias urbanas. Siendo estas descargadas de http://www.inec.gob.ec/estadisticas/?option=com_content&view=article&id=149 para obtener el DPA actualizada a Diciembre del año 2010

De esta manera, se obtuvo la información básica para la ubicación espacial del resto de geo datos en el cantón Quito (DMQ) y sus respectivas parroquias y zonas censales. (Ostle, 1974, pág. 63)

Cabe indicar que el INEC no incluye en su base de datos `dpa_inec.xls`, ni en los mapas `shapefile` y `jpeg`,

los códigos DPA de las parroquias urbanas, debido a que no es de competencia de ésta institución. La información de la división parroquial urbana la maneja cada municipio. Por esta razón fue necesaria la edición del archivo de Excel en el formato requerido para el proyecto, ya que se establecerán relaciones de validaciones dentro de los procesos de modelación espacial.

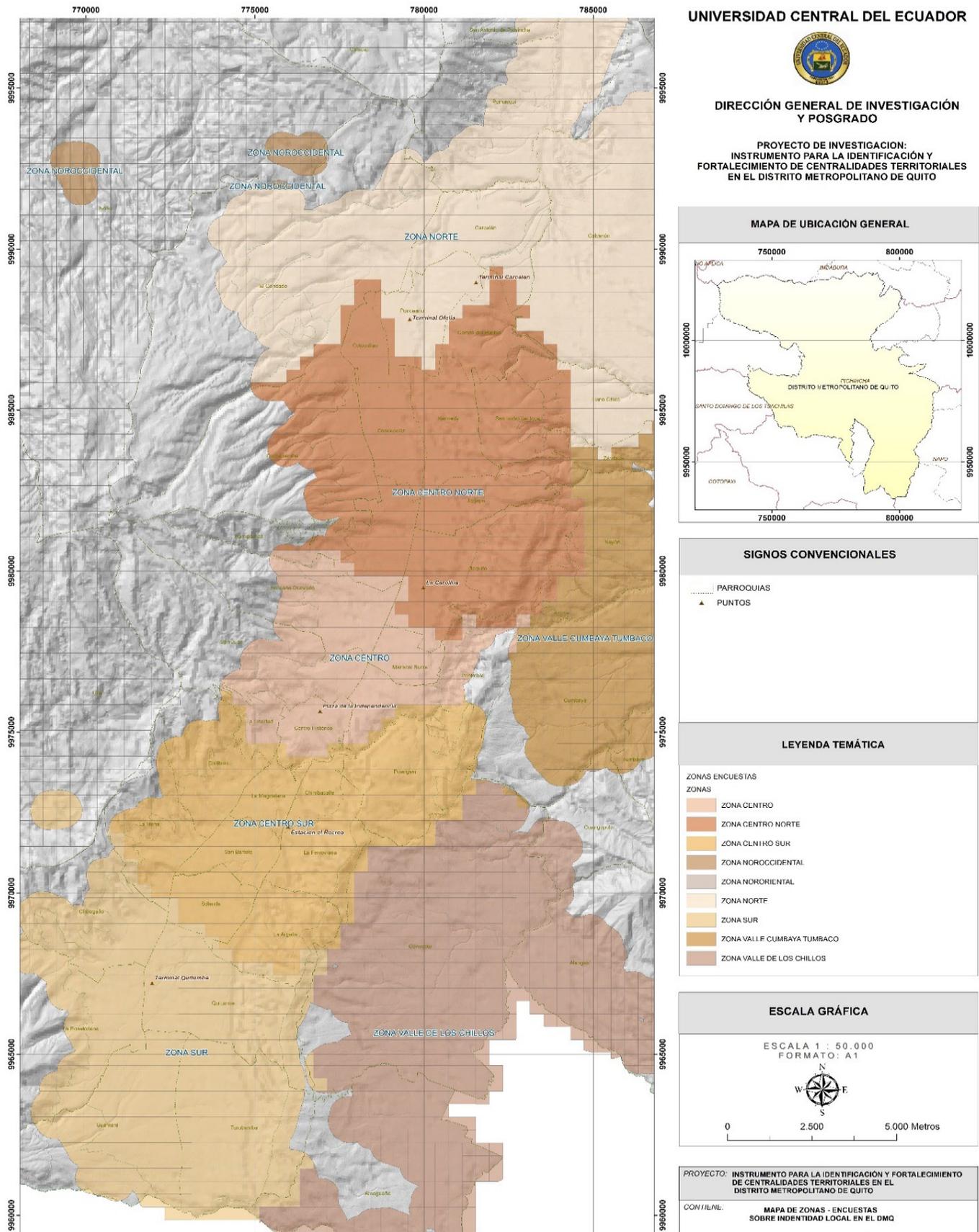
Se recopiló la mayoría de información geográfica de los geo datos básicos y sociales públicas, la mismas que disponen de enlaces hacia los geo portales o bases de datos, como son Municipio de Quito a través de su portal www.datosabiertos.quito.gob.ec Instituto Geográfico Militar, SENPLADES, Ministerio de Educación y el Ministerio de Salud Pública.

Entre la información recopilada se tiene archivos geo referenciados de mapas temáticos, hidrografía, vías y caminos, información de salud y educación, límites distritales, etc. Obteniendo un total de 1 655 073 personas contempladas desde los 10 años en adelante. Con el total de población se aplicó el cálculo del muestreo estratificado con afijación proporcional al tamaño.

La disponibilidad de información sobre la distribución de la población a gran escala, e independiente de las unidades administrativas de recogida de la información, es una necesidad en muchos ámbitos: evaluación de riesgos de terremotos, inundaciones o incendios, integración de estadísticas demográficas con información medioambiental en formatos diversos, diseño de políticas de prestación de servicios públicos o estudios de accesibilidad y movilidad en las grandes ciudades. Por estas razones, la literatura ha explorado numerosos métodos de desagregación de la población, recogida a nivel municipal o de sección censal, en polígonos urbanos procedentes de los sistemas de información de coberturas y usos del suelo o celdas de una malla regular, permitiendo de esta forma a generación de capas raster de población, fáciles de combinar con información medioambiental y geográfica. (Gisbert, 2016)



Figura 1 Distrito Metropolitano de Quito y Zonas Censales Fuente: Dirección de información Cartográfica Estadística –DICE- a nivel de zonas censales del Distrito Metropolitano de Quito
Elaborado por: Ing. Gabriela Paz





Según (Scheaffer, Mendenhall, & Ott, pág. 81) mencionan que después de que las unidades de muestreo han sido seleccionados de una muestra aleatoria de cada estrato debemos estar seguros de que las muestras seleccionadas sean independientes. Esto, debe ser aplicando técnicas de muestreo aleatorio dentro de cada estrato, de tal manera que las observaciones elegidas en un estrato no dependan de las que se han elegido en otro estrato.

L = número de estrato

N_i = número de unidades muestrales en el estrato i

N = número de unidades muestrales en la población

$N = N_1 + N_2 + \dots + N_L$

Para el cálculo de la muestra se considera los siguientes criterios:

Para (Kish, 1972, pág. 66) Un buen diseño de muestreo requiere que se equilibren sensatamente cuatro criterios amplios:

Orientación hacia la meta

El diseño completo, tanto en la selección como la estimación, debe orientarse a los objetivos de la investigación, hechos a la medida del diseño de la encuesta y ajustados a las condiciones de la encuesta. Estas consideraciones deben influir en la selección y en la definición de la población, en la medición y en los procedimientos de muestreo.

La medibilidad

Permite calcular, a partir de la propia muestra, estimaciones válidas o aproximaciones de su variabilidad de muestreo.

Esta es la base necesaria para la inferencia estadística, y sirve como puente científico y objetivo, entre el resultado de la muestra y el valor desconocido de la población.

La practicidad

Se refiere a los problemas que deben resolverse para llevar a cabo el diseño esencialmente como se propuso. Una muestra probabilística no puede crearse por suposición, ni está dada, como sucede en los problemas teóricos.

La simplicidad debe estar siempre entre los objetivos, porque reduce el riesgo de los errores, y compensa con creces la pérdida de un poco de eficiencia teórica.

Una buena parte del trabajo práctico consiste en

suponer qué y cuántas irregularidades se podrán tolerar y en dónde serán admisibles.

La economía

Se refiere a cumplir los objetivos de la encuesta con un costo (esfuerzo) mínimo y el grado en el que se alcanza este objetivo.

La eficiencia mide el número de elementos de la muestra que se requieren para una precisión fija. Una muestra será demasiado pequeña si sus resultados carecen de la precisión suficiente para contribuir de manera apreciable en las decisiones.

Para determinar el tamaño de la muestra según Triola (2006, pág. 299) Usamos como el estimado puntual de p , al no ser sesgado es decir: la distribución de las proporciones muestrales tienen al centro para el valor de p . La proporción muestral es el estimador más consistente en el sentido de que la desviación estándar de la proporción muestral tiende a ser menor que la desviación estándar de cualquier otro estimador sin sesgo.

Para determinar el tamaño da la muestra se deben tomar en cuenta algunos aspectos, que se relacione con: el parámetro, estimador, el sesgo, el error muestral, el nivel de confianza y la varianza poblacional. (Kish, 1972, págs. 73-77)

El parámetro se refiere a la característica de la población que es objeto de estudio y el estimador es la función de la muestra que se usa para medirlo.

El tamaño de muestra se definió teniendo en cuenta:

Nivel de confianza y la precisión

Errores relativos del muestreo

La no respuesta o no entrevistas (rechazos ausente y problemas de marco)

El peso que se le dio a cada una de las celdas, está definido según la definición de muestreo por estrato con afijación proporcional al tamaño. (Gutiérrez y de la Vara & Namakforoosh, 2008) Indican que la decisión para la determinación del tamaño muestral dependerá de varios factores:

La magnitud de las diferencias que se quiere detectar en la investigación, es decir, la importancia de la decisión a tomar. Si son pequeñas mayor será el tamaño muestral y viceversa.

La variación esperada en los datos, debido a fuentes de variación no controladas. A mayor variación será

necesario un tamaño muestral mayor.

El número de tratamientos (o muestras) que se desea comparar. A mayor número de ellos, menor tamaño muestral.

Riesgo que está dispuesto a tomar el investigador. A menor riesgo, el tamaño deberá ser mayor. Aquí está incluida la potencia de la prueba que se desea.

La complejidad de los análisis estadísticos. Cuanto más complejos el tamaño deberá ser más grande.

El número de variables o factores de estudio. Cuanto más numerosas, más grande tendrá que ser la muestra.

Para poblaciones finitas, el tamaño de muestra será menor que para poblaciones infinitas, determinando una relación no lineal, debido a la significancia y al error aplicado en el cálculo. Mateu y Casal (2003, págs. 1:8-14)

MUESTRA PARA POBLACIONES FINITAS

Si el tamaño de la población o universo es conocido, la elección de la muestra, está basada a partir de ello, donde el número de elementos de la muestra puede llegar a ser una proporción considerable de la población. En tal situación, puede entenderse fácilmente que la precisión de la estimación sería superior, al estar mejor representada al conjunto de la población. (Spiegel, M. R., & Stephens, L. J, 2002).

En concreto, para el caso de poblaciones finitas, puede demostrarse que la varianza del estimador propia de poblaciones infinitas, se debe corregir por un factor. Este factor de corrección sería $(N-n/N-1)$.

Para la determinación del tamaño de la muestra se utilizó la siguiente expresión:

$$n = \frac{z^2 \cdot p \cdot q}{e^2} \quad \text{Ec. (1)}$$

Personas

Donde:

N= población o universo

z= coeficiente de confianza (1,96)

p= éxito

q= 1-p

e= error relativo máximo esperado

n=tamaño de la muestra

Desviación del error $\pm 3 \sigma$

CÁLCULOS

DATOS	
N	1655073
P	0,5
Q	0,5
E	1%
Z	1,96

El nivel de confianza, es la probabilidad $1-\alpha$ de estimación efectuada se ajuste a la realidad; es decir, que caiga dentro de un intervalo determinado basado en el estimador y que capte el valor verdadero del parámetro a medir. (Mateu y Casal, 2003, pág. 301)

El error muestral, es la diferencia máxima probable (con probabilidad $1-\alpha$) entre la proporción muestral que se observa y el valor real de la proporción de la población. Es importante que los parámetros determinados para este coeficiente de confianza como para el error relativo no superen los establecidos por la literatura estadística. En este sentido, se decidió que el coeficiente de confianza sea el de la distribución normal que garantiza el 95% de confianza, y el error relativo no supere el 10%.

Para la investigación se toma en cuenta que los datos obtenidos son necesarios para aplicarlos gráficamente en un mapa, por lo que se tiene el 95% de confianza y el 1% de error con el único fin de poderlo graficar, para estudios de Urbanismo e Identificación territorial. (Mateu y Casal, 2003, pág. 305) Se utiliza la máxima variabilidad, en el caso de no existir antecedentes sobre la investigación (no hay otras o no se pudo aplicar una prueba previa), entonces los valores de variabilidad es $p=q=0.5$.

Se utilizó la variabilidad máximo donde $p \cdot q = 0.5 \cdot 0.5$ ya que no se tiene información de la varianza y es más eficiente tomar la fórmula donde nuestra información está basada en probabilidades.

EL TAMAÑO MUESTRAL AJUSTADO A LAS PÉRDIDAS

Al hablar del levantamiento de información de aproximadamente de 9549 encuestas y a estas ser procesadas en una base de datos Google Forms, la misma que trabaja de forma on-line permitiendo recopilar toda la información levantada por los encuestadores de forma rápida, desde cualquier lugar y a cualquier hora actualizando la base de datos en tiempo real. Por tal motivo de prever que en el proceso exista pérdida de información, abandono o no respuesta.

En todos los estudios es preciso estimar pérdidas de información por diversas razones (pérdida de información, abandono, no respuesta) por lo que se debe



incrementar el tamaño muestral respecto a dichas pérdidas. (Mateu y Casal, 2003, pág. 308)

TASA DE NO RESPUESTA

El patrón general de tasas de no respuesta más altas al principio del proceso en cuestión y más bajas al final en general no se modifica ni en las entrevistas domiciliarias. Cuando no existe información o investigaciones similares, la tasa de no respuesta se fija como un porcentaje aproximado, utilizada como una hipótesis.

Se determina una tasa de no respuesta, por lo cual para el caso de estudio utilizamos la siguiente fórmula:

Tasa de no respuesta (TNR)=

Donde:

n: Número de sujetos sin pérdidas

R: Proporción esperada de pérdidas.

Según estudios el INEC la proporción de no respuesta aceptable para un estudio de investigación se encuentra alrededor del 9,1%(INEC,2016).

Cálculo:

n: 9548

R: 9.1%

TNR=

TNR= 9548 *

TNR=10504

Incluyendo el porcentaje de la tasa de no respuesta, la muestra a levantar para el estudio es de 10504 personas en el DMQ.

MARCO MUESTRAL

El marco muestral está considerado por toda el área del Distrito Metropolitano de Quito el cual está dividido en áreas de 400 m por 400m (celdas) considerando la distancia que una persona está dispuesta a caminar antes de tomar un vehículo o transporte público según (ESquivel-Cuevas, Hernández Mercado, & Garnica Mornroy, 2013, págs. 24,25) Para determinar los rangos con los cuales se fijaría la ponderación de los diferentes indicadores se realizó un estudio para determinar la distancia que un peatón estaba dispuesto a caminar en un entorno urbano dado, usando una aplicación para teléfonos celulares con GPS integrado llamada "Everytrail", la cual, registra distancia, tiempo, velocidad promedio y pendiente durante un recorrido. Los rangos resultantes fueron usados para

evaluar el nivel de cobertura de un equipamiento o servicio y representan el grado de accesibilidad peatonal desde el punto de origen, obteniendo mayor ponderación a mayor cercanía. Se pueden interpretar en términos de distancia o de tiempo de la siguiente manera:

5 minutos = 0-400 metros, 10 minutos = 400-800 metros, 20 minutos = 800-1600 metros y 30 minutos =1600-2400 metros.

La distancia máxima recomendada a la estación más cercana de transporte público masivo para un desarrollo orientado al transporte es de 1 kilómetro, que presentan una caminata de 15 a 20 minutos. Al construir mayor densidad más cerca de las estaciones de transporte, un desarrollo puede maximizar el número de personas y servicios a los que se puede acceder caminando. Institute for Transportation and Development Polic (2003, pág. 23). El diseño muestral de la encuesta se caracteriza por ser probabilístico, en consecuencia, los resultados obtenidos se pueden generalizarse a toda la población. El diseño a su vez, es estratificado y proporcional al tamaño de la población, mejorando la eficiencia del diseño.

En la investigación se tomó como intervalo de tiempo 5 minutos igual a 400 metros de distancia que es un equivalen a cuatro manzanas tradicionales (tales como las del Centro Histórico o el barrio de La Mariscal de Quito) lo que nos permite determinar el número de estrato y codificarlos para establecer la ponderación de los mismos según la densidad poblacional, obteniendo 5571 celdas en todo la parte urbana del DMQ, para la aplicación de encuestas se consideró la eliminación de las celdas donde existía áreas verdes, parques, estadios, quebradas etc. dando un total de 2824 celdas censales.

Según Mariani, (2013, pág 5) la densidad de población, también denominada a veces formalmente población relativa (para diferenciarla de la absoluta, la cual simplemente equivale a una determinada cantidad de habitantes), se refiere al número promedio de habitantes de un área urbana o rural en relación a una unidad de superficie dada.

Medimos la densidad de una ciudad en Habitantes por hectárea, en América Latina (y en todas partes del mundo en donde se utiliza el sistema métrico decimal), también se puede hablar con respecto a una ciudad completa o territorio, habitantes por kilómetro cuadrado y en casos específicos, habitantes por metro cuadrado.

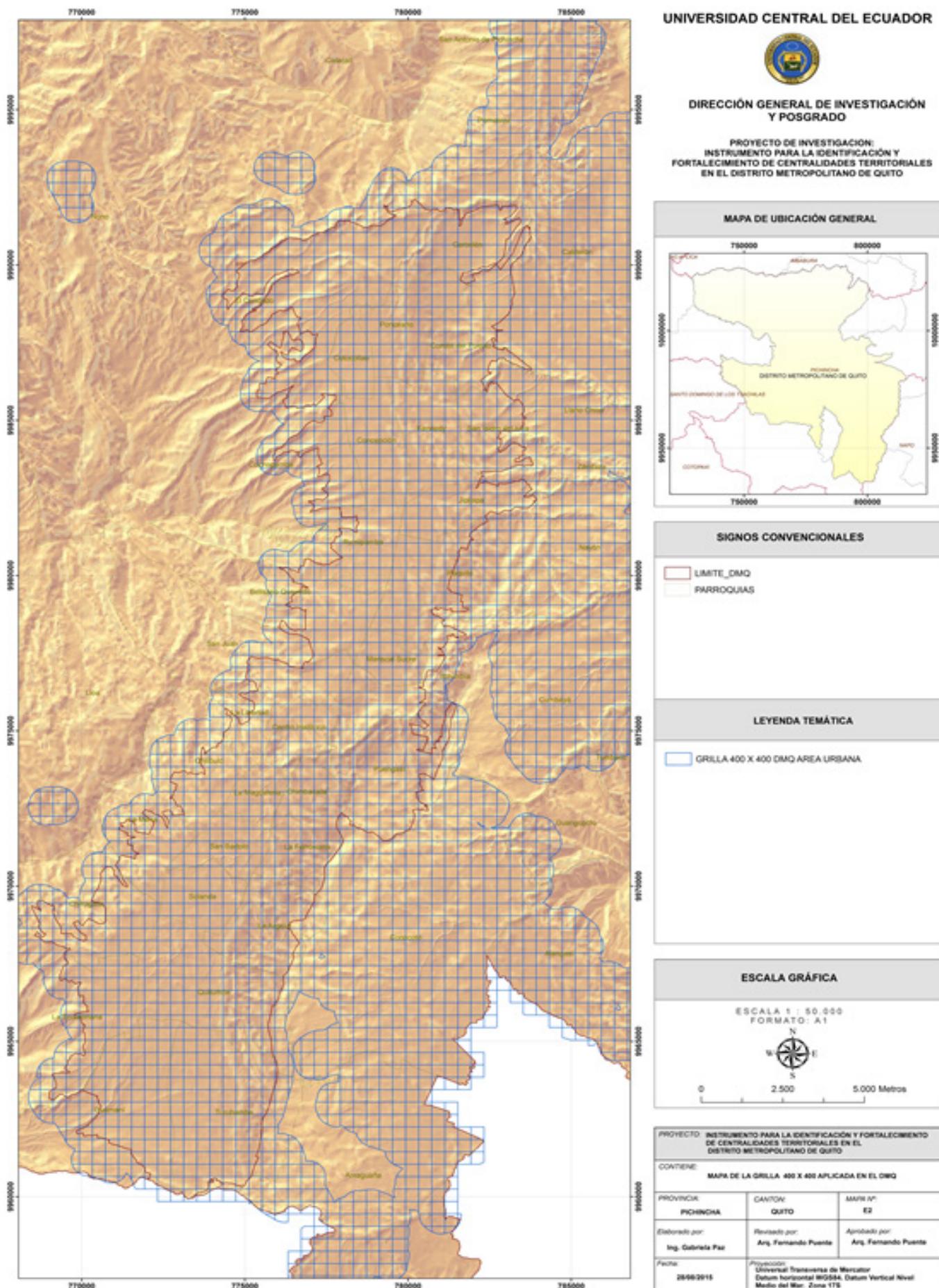


Figura 2: Mapa de distribución de las encuestas en el Distrito Metropolitano de Quito por celdas 400*400 metros. Fuente: INEC 2010 Elaborado por: Ing. Henry Vilatuña e Ing. Gabriela Paz



MUESTREO ESTRATIFICADO

Para realizar muestras para poblaciones finitas, teniendo en cuenta que se conoce la ubicación geográfica y la distribución de la población, es mejor aplicar un muestro estratificado el cual consiste en los siguientes pasos: 1) La población completa de unidades de muestreo se divide en subpoblaciones distintas llamadas estrato. 2) Dentro de cada estrato se selecciona una muestra separada a partir de todas las unidades de muestreo que componen este estrato. 3) De la muestra obtenida de cada estrato, se calcula una media separada (u otra estadística). Estas medidas por estrato se ponderan adecuadamente para formar una estimación combinada de la población completa.

4) las varianzas también se calculan separadamente dentro de cada estrato y entonces se ponderan adecuadamente y se suman en una estimación combinada para la población. (Kish, 1972, pág. 101)

Hay tres razones por las que comúnmente se recurre a muestreo estratificado:

El tamaño de los estratos permiten obtener una muestra de acuerdo al peso que cada uno representa, permitiendo disminuir las varianzas de las estimaciones de la muestra.

Se pueden formar los estratos para utilizar diferentes métodos y procedimientos dentro de ellos.

Los estratos pueden establecerse porque las subpoblaciones dentro de ellos también se definen como dominios de estudio. Un dominio es una parte de la población para la que se han planeado estimaciones separadas en el diseño de la muestra. Se llama dominio a cualquier subdivisión acerca de la cual se planea la encuesta para proporcionar información numérica de precisión conocida.

El método de muestreo por estrato fue seleccionado porque al existir diferenciación entre las celdas es necesario obtener una muestra representativa de cada una de ellas, siendo éstas los estrato de las cuales se obtuvieron los pesos relativos, obteniendo la muestra con afijación proporcional al tamaño, dándonos a conocer el número de personas a las cuales se van a encuestar. (Kish, 1972, pág. 105) Para el cálculo del número de encuestas por estrato con afijación proporcional al tamaño se utiliza la siguiente fórmula:

ESTRATO

Ec. (2)

Donde:

i =es el número de estrato

N = tamaño de la población dentro de cada estrato

Y_i = peso de la población dentro de ese estrato.

Para la elaboración de la muestra, se obtuvo 5572 estrato (celdas) de los cuales 2746 fueron segregados por ser áreas verdes, parques, estadios, quebradas, centros de estudio donde la población no es representativa, por lo que no se llega a realizar ni una encuesta, tomamos en cuenta sólo 2824 estrato (celdas) para ser encuestadas, considerando que cuentan con población suficiente para hacer por lo menos una encuesta.

Para demostración, procedemos a juntar las celdas en grupos de 250 y una última con 74 celdas, donde observamos que el mínimo de encuestas a realizarse es de 58 y un máximo de 1583.

El cálculo de la muestra está realizado con respecto a la ponderación al tamaño poblacional previamente calculado en un Sistema de Información Geográfica (SIG), lo que nos permite dejar un lado, los sectores donde no existe población. Un muestreo en el que haya estratificación y proporcionalidad, además de selección al azar, nos daría mayor seguridad de que la muestra sea representativa, pero de todos modos, cualquier tipo de muestreo probabilístico nos permite tener confianza sobre la representatividad de la muestra. Además, hay que tener en cuenta que “la estratificación requiere un mayor esfuerzo y un mayor tiempo que la simple aleatorización”, (Pereda Marín, S, 1987 pág. 130).

Dentro de la asignación del número de encuestas para cada estrato o celda se consideró intervalos de 80 -140 habitantes por celda aplicando redondeo para obtener el número total de habitantes dentro de la muestra.

RESULTADOS

Cálculo de la población contemplada desde los 10 años de edad

La estimación del tamaño de población se generó a partir del censo 2010, de la cual se segregó la Provincia de Pichincha, para luego tomar las parroquias Urbanas y Rurales del Distrito Metropolitano de Quito, así como las personas que tengan de 10 años en adelante, de los cuales obtuvimos 1 655 073 personas.

Tabla 1. Población del DMQ

DISTRITO METROPOLITANO DE QUITO	(Área sub polígono)	Población en el sub-polígono	Total
Total general	77102,20265	1655072,8	1655073

Elaborado por: Mayra Díaz

Fuente: Censo INEC 2010

Tabla 2. Distribución de la población DMQ por edad

Edad	Población
De 10_15	199637
De 15_20	199568
De 20_25	210549
De 25_30	205976
De 30_35	179526
De 35_40	155493
De 40_45	133435
De 45_50	104897
De 50_55	100434
De 55_60	82640
De 60_65	62918
Total	1655073

Elaborado por: Mayra Díaz

Fuente: Polígonos Censales/ CENSO 2010/ INEC

El cálculo aproximado de la población en base a los polígonos censales a partir de los 10 años de edad y con amplitud de 5 años.

Mediante este tamaño de población se procede al cálculo de la muestra.

Tabla 3. Resultados del cálculo de la muestra

Grupos de celdas	CODIFICA_2	Población	Frecuencia porcentual	N peso estratificada	Muestra
Grupo1	250 celdas	74814	0,04520284	474,7654285	522
Grupo2	250 celdas	131849	0,079663556	836,7063247	858
Grupo3	250 celdas	139380	0,084213808	884,4976264	905
Grupo4	250 celdas	247487	0,149532377	1570,538557	1583
Grupo5	250 celdas	210068	0,106923707	1333,079691	1347
Grupo6	250 celdas	118651	0,071689285	752,9525604	760
Grupo7	250 celdas	97657	0,059004648	619,7258193	653

Elaborado por: Mayra Díaz

Fuente: Polígonos Censales/ CENSO 2010/ INEC

Del total de celdas extendidas dentro del DMQ se formó diez grupos contemplado de 250 celdas cada

una para efecto de levantamiento de información la misma que contempla la población desde 10 años hasta los 65 años, el porcentaje de cada grupo, el peso estratificado para el peso de la muestra y el total de la muestra a ser aplicado para el levantamiento de la información.

Tabla 4. Número de encuestas a aplicar de acuerdo a la celda y en función de la población.

Intervalos		N° Celdas
Límite inferior	Límite superior	
1	9	2566
9	17	229
17	25	27
25	33	2
Total		2824

Elaborado por: Mayra Díaz

Fuente: Polígonos Censales/ CENSO 2010/ INEC



Existe un total de 2566 celdas en las que se aplica de 1 a 9 encuestas para el levantamiento de información. Así como también existe 2 celdas donde la población es grande por lo que se realiza de 25 a 33 encuestas y estas se distribuyen en todo el DMQ de la siguiente manera:

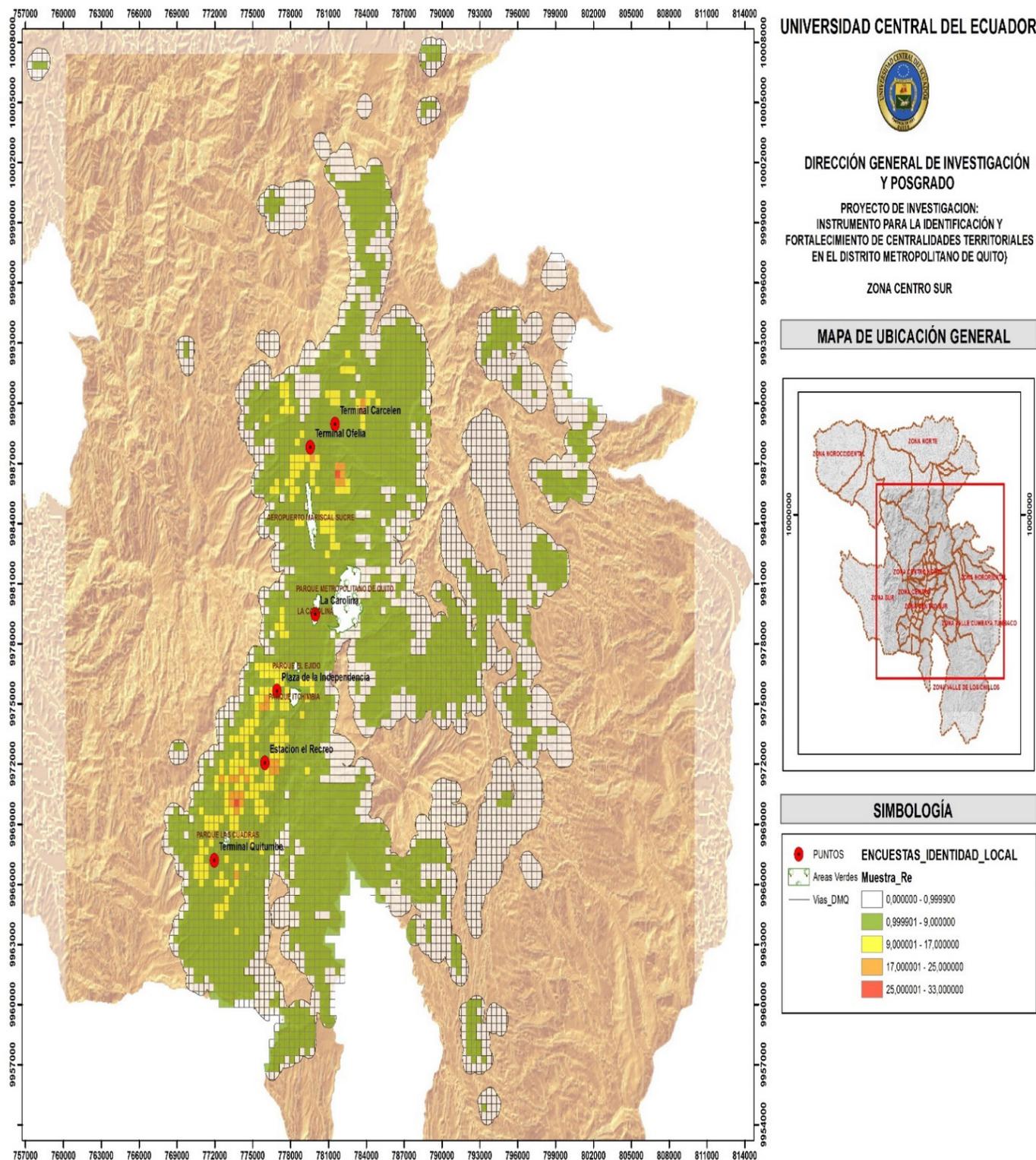


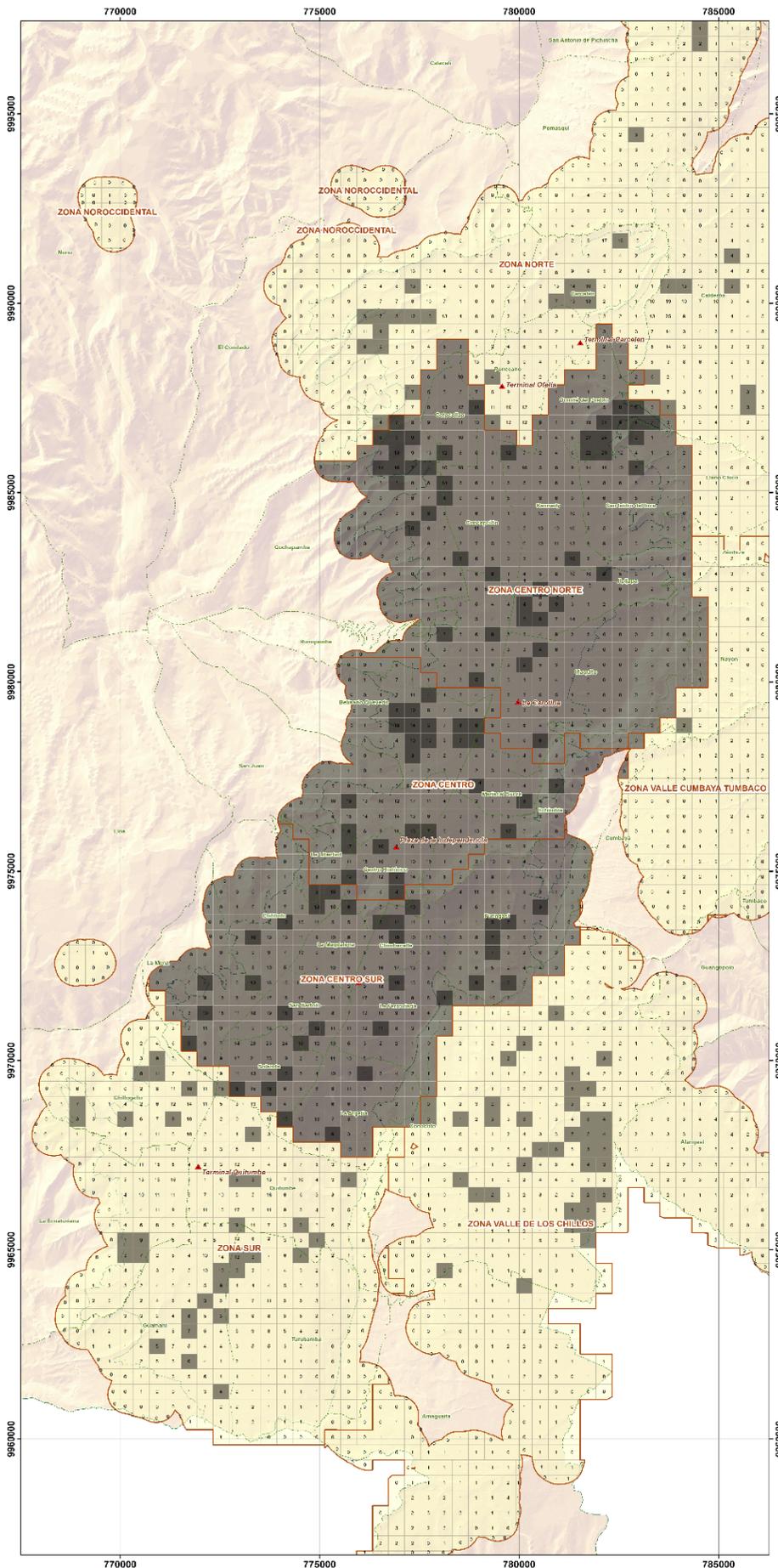
Figura 3. Mapa de distribución de las encuestas en el Distrito Metropolitano de Quito por celdas 400*400 metros Fuente: Dirección de información Cartográfica Estadística –DICE- a nivel de zonas

Elaborado por: Ing.Henry Vilatuña e Ing.Gabriela Paz

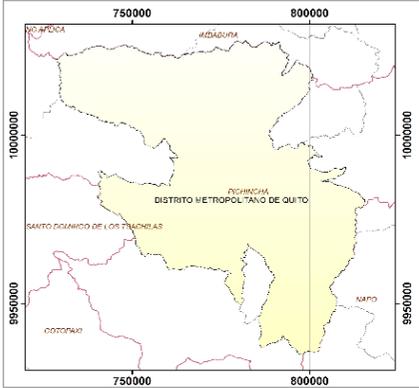


DIRECCIÓN GENERAL DE INVESTIGACIÓN Y POSGRADO

PROYECTO DE INVESTIGACIÓN:
INSTRUMENTO PARA LA IDENTIFICACIÓN Y FORTALECIMIENTO DE CENTRALIDADES TERRITORIALES EN EL DISTRITO METROPOLITANO DE QUITO



MAPA DE UBICACIÓN GENERAL



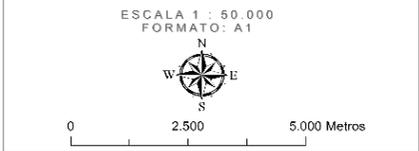
SIGNOS CONVENCIONALES

- ▲ PUNTOS
- ▭ PARROQUIAS
- ▭ LIMITE ZONAS ENCUESTAS

LEYENDA TEMÁTICA

- ZONAS ENCUESTAS
- ZONA CENTRO
 - ZONA CENTRO NORTE
 - ZONA CENTRO SUR
 - ZONA NOROCCIDENTAL
 - ZONA NORORIENTAL
 - ZONA NORTE
 - ZONA SUR
 - ZONA VALLE CUMBAYA TUMBACO
 - ZONA VALLE DE LOS CHILLOS

ESCALA GRÁFICA



PROYECTO: INSTRUMENTO PARA LA IDENTIFICACIÓN Y FORTALECIMIENTO DE CENTRALIDADES TERRITORIALES EN EL DISTRITO METROPOLITANO DE QUITO

CONTIENE: MAPA DE ENCUESTAS TOMADAS SOBRE IDENTIDAD LOCAL EN EL DMQ

PROVINCIA: PICHINCHA	CANTON: QUITO	MAPA N°: E1
Elaborado por: Ing. Gabriela Paz	Revisado por: Arq. Diego Hurtado	Aprobado por: Arq. Diego Hurtado
Fecha: 01/04/2016	Proyección: Universal Transversa de Mercator Datum horizontal WGS84, Datum Vertical Nivel Medio del Mar. Zona ITS	

Figura 4. Mapa de distribución del número encuestas por celda en el Distrito Metropolitano de Quito / Fuente: Dirección de información Cartográfica Estadística –DICE- a nivel de zonas. / Elaborado por: Ing. Henry Vilatuña e Ing. Gabriela Paz



CONCLUSIONES

Aproximadamente existe 1 655 073 personas contempladas desde los 10 años hasta los 65 años de edad dentro del DMQ, la base de datos con la se trabajó para la obtención fueron: el censo del INEC 2010, Dirección de información Cartográfica Estadística – DICE- a nivel de zonas censales del Distrito Metropolitano de Quito.

En base de la población obtenida y con aplicación del muestreo estratificado con afijación proporcional, se obtuvo los pesos para cada grupo de trabajo conformadas por 250 celdas donde se evidencia: la población, el porcentaje y el número de encuestas a ser aplicadas en función de la población.

De un total de 5571 celdas de toda la parte urbana del DMQ solo se consideró 2824 celdas para la aplicación de encuestas y de ellas 29 celdas es donde se encuentra la mayor concentración poblacional para lo cual se aplica un promedio de 25 encuestas por celda con un intervalo de 80-140 habitantes por encuesta aproximadamente.

Los sectores donde se concentra la mayor población dentro de las celdas son: Comité del Pueblo, Cotocollao, Centro Histórico, La Magdalena, Chimbacalle, San Bartolo, Ferroviaria y Solanda son los sectores donde se aplican la mayor cantidad de encuestas de entre 10 a 33 encuestas tomando en consideración el intervalo de 80 a 140 habitantes por celda.

BIBLIOGRAFÍA

- [1] Argibay, J. C. (2009). Muestra en investigación cuantitativa. Subjetividad y procesos cognitivos, http://www.scielo.org.ar/scielo.php?script=sci_abstract&pid=S1852-73102009000100001&lng=es&nrm=iso&tlng=es
- [2] Bravo, D. N. (2001). Jerarquización urbana funcional y análisis de los patrones de distribución espacial del índice de atracción de los núcleos de población de La Rioja. Berceo, (141), 183–194. <https://dialnet.unirioja.es/servlet/articulo?codigo=224571>
- [3] Esquivel-Cuevas, M., Hernández Mercado, O. A., & Garnica Monrroy, R. (06 de Agosto de 2013). Modelo de Accesibilidad Peatonal (MAP). Bitácora de la Universidad de Colombia-Bogotá, 30.
- [4] Gisbert, F. J. G. (2016). Una aproximación volumétrica a la desagregación espacial de la población combinando cartografía temática y datos LIDAR. Working papers = Documentos de trabajo: Serie EC (Instituto Valenciano de Investigaciones Económicas), (3), 1. Recuperado a partir de <https://dialnet.unirioja.es/servlet/articulo?codigo=5617872>
- [5] Gutiérrez y de la Vara, & Namakforoosh. (2008). Tamaño de muestras en Estudios Clínicos. Institute for Transportation and Development Polic. (2003). Tod Standard. Investigación, New York.
- [6] Kish, L. (1972). Muestreo de encuestas. (R. Cruz López, Trad.) México: Trillas.
- [7] Mariani, G. (19 de Marzo de 2013). Las dimensiones del Urbanismo: espacio, densidad, tiempo y distancia. La ciudad viva, 35.
- [8] Casal, J, Mateu,E (2003).Tamaño de la Muestra Universitat Autònoma de Barcelona, http://optimierung.mathematik.uni-kl.de/mamaeusch/veroeffentlichungen/ver_texte/sampling_es.pdf
- [9] Murray, S. (1988). Estadística (2º ed.). Madrid: McGraw Hill.
- [10] Ostle, B. (1974). Técnicas de la Estadística Moderna, Cuando y dónde aplicarlas (cuarta ed.). México, México: Editorial Limusa.
- [11] Scheaffer, R. L., Mendenhall, W., & Ott, L. (s.f.). Elementos de muestreo. (G. Rendón Sánchez, & J. R. Gómez Aguilar, Trad.) Mexico, Chapingo, Mexico: Grupo Editorial Iberoamericano.
- [12] SISTEMA DE BIBLIOTECAS EPN - Koha › Detalles para: Fortalecimiento de centralidades urbanas de Quito. (s. f.). http://biblioteca.epn.edu.ec/cgi-bin/koha/opacdetail.pl?biblionumber=23514&shelfbrowse_itemnumber=31169
- [13] Spiegel, M. R., & Stephens, L. J. (2002). Estadística (Tercera ed.). (L. E. Pineda Ayala, Trad.) México, México: McGraw-Hill Interamericana.
- [14] Triola, M. F. (2006). Estadística (Novena ed.). (P. M. Guerrero Rosas, Ed.) México, México: Pearson.
- [15] Talavera, J. O., & Rivas-Ruiz, R. (2013). IV. Pertinencia de la prueba estadística. IV. Relevance of Statistical Tests., 51, S30–S34. <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=97004210&lang=es&site=ehost-live>
- [16] Unidas, N. (2009). Diseño de muestras para encuestas de hogares Directrices prácticas. New York.