

Formalización del mercado laboral en Cali

Propuesta para el Ministerio de Trabajo

Diseño muestral

En esta sección se describe la población de interés, se definen las unidades muestrales y se especifica el diseño muestral propuesto. Aquí queda definido el mecanismo para la selección de los hogares y las personas a los que se aplicará la encuesta sobre empleo y calidad de vida en Cali.

1. Universo de estudio

Todos los hogares residentes en la ciudad de Cali; incluye veintidós (22) comunas en el área urbana y una (1) zona ubicada en el área rural del municipio. En la tabla 1 se presenta el número de hogares en el universo de estudio por zona:

Tabla 1. Número de hogares por comuna en Cali

Zona	Número de hogares	% número de hogares
Total Cali	616.775	100%
Comuna 1	22.370	3,6%
Comuna 2	29.967	4,9%
Comuna 3	12.480	2,0%
Comuna 4	14.597	2,4%
Comuna 5	29.789	4,8%
Comuna 6	50.378	8,2%
Comuna 7	19.557	3,2%
Comuna 8	27.751	4,5%
Comuna 9	12.355	2,0%
Comuna 10	29.720	4,8%
Comuna 11	28.682	4,7%
Comuna 12	18.167	2,9%
Comuna 13	47.791	7,7%
Comuna 14	45.700	7,4%
Comuna 15	41.390	6,7%
Comuna 16	28.409	4,6%
Comuna 17	36.004	5,8%
Comuna 18	30.465	4,9%

Zona	Número de hogares	% número de hogares
Comuna 19	30.013	4,9%
Comuna 20	18.549	3,0%
Comuna 21	29.259	4,7%
Comuna 22	2.859	0,5%
Rural	10.523	1,7%

Fuente: elaboración propia a partir de las proyecciones de población a 2012 del DANE

Como información adicional, de acuerdo a las proyecciones de población a 2012 que realiza en DANE, Cali presenta 2.29 millones de habitantes; el 27% de la población de Cali es Afro-descendiente (incluye: Negro, mulato y afro colombiano según el informe del DANE).

El marco muestral disponible en el Centro Nacional de Consultoría lo constituye el listado de viviendas, hogares y personas (VIHOPE) del Censo Nacional de Población de 2005 realizado por el DANE. Esta herramienta permite la selección de manzanas, viviendas y personas que requiere el Diseño.

2. Unidades muestrales

Unidad de observación: La unidad estadística de observación es el hogar. El respondiente idóneo es una persona de 18 años o más, diferente al personal de servicio, quien brinda la información de todas las personas que componen el hogar

Unidades de muestreo: de acuerdo al diseño, se tienen 2 unidades de muestreo: (1) Manzana y (2) hogar.

3. Diseño muestral

La estrategia muestral planteada para la realización del estudio es un diseño Probabilístico, estratificado en 2 etapas, con selección de unidades por Muestreo Aleatorio Simple en cada etapa. Su correcta ejecución garantiza que el mecanismo de selección de la muestra y la proyección de resultados cumplen con la precisión y calidad requeridas.

De acuerdo a los requerimientos técnicos del estudio, el universo de estudio se estratificará por la variable geográfica “**comuna**”. De esta manera se tienen 23 estratos (22 en el área urbana y 1 en el área rural). A continuación el detalle de la selección de unidades de muestreo en cada estrato, siguiendo el proceso en dos etapas:

Etapas I. Selección de manzanas

La selección de manzanas dentro de cada una de las comunas en el área urbana se realizará por Muestreo Aleatorio Simple. Se visitarán todos los barrios en la ciudad de Cali; el mecanismo de selección llamado “coordinado negativo” permitirá la selección de manzanas dentro de cada uno de los barrios, cumpliendo así con el requerimiento de que el operativo de campo tenga cobertura del 100% de los barrios. En el área rural, se tomarán todas las veredas.

Método coordinado negativo: Un algoritmo en este método es el siguiente:

- i) para cada manzana en el barrio objetivo, seleccione un número aleatorio entre cero y uno en Excel
- ii) fije los número aleatorios como valores
- iii) ordene la lista de manzanas por el número aleatorio (de mayor a menor)
- iv) tome la cantidad de manzanas definida previamente, seleccionando las primeras de la lista.

Etapas II. Selección de hogares dentro de cada manzana seleccionada

La selección de hogares dentro de cada manzana se realizará por Muestreo Aleatorio Simple. De acuerdo a las prácticas comúnmente aceptadas en las investigaciones en hogares en Colombia, se seleccionará aleatoriamente una vivienda por cada lado de la manzana. De manera idéntica, en el área rural se seleccionarán viviendas dentro de la vereda. En la vivienda seleccionada, en el caso de que se encuentre más de un hogar, se seleccionará el hogar de la primera persona idónea que se contacte para responder la encuesta.

Tamaño de muestra:

El tamaño de muestra propuesto dentro del diseño planteado es de **8.600** hogares. Este tamaño de muestra y el diseño de muestreo planteado garantizan un error de muestreo esperado no superior a 1.2% para el Total Cali y no superior a 7% para los resultados en cada una de las comunas. De hecho, en las comunas más grandes se presentarán resultados con errores de muestreo cercanos a 5%. Los cálculos realizados consideran una confiabilidad del 95% y un efecto de diseño de 1.3.

Nota 1. El efecto de diseño es un indicador que compara la varianza de un estimador del parámetro de interés bajo un diseño de muestreo cualquiera y la varianza del estimador bajo el conocido diseño MAS (Muestreo Aleatorio Simple); acá se está considerado que la variabilidad asociada al estimador en el diseño propuesto (estratificado en tres etapas) puede ser hasta un 30% mayor que la asociada al estimador si se realizara Muestreo Aleatorio Simple (sin estratificación ni etapas).

La tabla 2 presenta la propuesta de tamaño de muestra por comuna e incluye el error de muestreo máximo por comuna.

Tabla 2. Tamaño de muestra y error de muestreo máximo por comuna

Zona	Número de hogares	Tamaño de muestra	Error muestral máximo
Total Cali	616.775	8600	1,2%
Comuna 1	22.370	300	6,6%
Comuna 2	29.967	400	5,7%
Comuna 3	12.480	250	7,0%
Comuna 4	14.597	250	7,0%
Comuna 5	29.789	400	5,7%
Comuna 6	50.378	600	4,6%
Comuna 7	19.557	260	7,0%
Comuna 8	27.751	370	5,9%
Comuna 9	12.355	250	7,0%
Comuna 10	29.720	400	5,7%
Comuna 11	28.682	400	5,7%
Comuna 12	18.167	250	7,0%
Comuna 13	47.791	550	4,9%
Comuna 14	45.700	500	5,1%
Comuna 15	41.390	520	5,0%
Comuna 16	28.409	400	5,7%
Comuna 17	36.004	450	5,4%
Comuna 18	30.465	450	5,4%
Comuna 19	30.013	450	5,4%
Comuna 20	18.549	250	7,0%
Comuna 21	29.259	400	5,7%
Comuna 22	2.859	250	7,0%
Rural	10.523	250	7,0%

Aplicando los conceptos teóricos en muestreo al caso de este estudio, el segmento de población afrocolombianas en Cali queda automáticamente representado en la muestra si se siguen los procedimientos probabilísticos de selección ya descritos anteriormente; aún más, como la selección se realizará por Muestreo Aleatorio Simple, la proporción de personas afrodescendientes en la muestra será cercana al 27% (porcentaje de afrodescendientes en la población en estudio), en todo caso, el Centro Nacional de Consultoría garantiza que el tamaño de muestra para el segmento de población

afrodescendiente será suficiente para garantizar el error de muestreo requerido. En el caso de los niveles socioeconómicos, el control se realizará por microrepresentatividad.

La fórmula utilizada para el cálculo del tamaño de la muestra es la siguiente:

$$n = \frac{Np(1-p)(Deff)}{N(Esrel * p)^2 + (p(1-p)(Deff))}$$

Dónde,

N → Tamaño del universo

p → Proporción esperada como estimación de la variable principal en estudio

Esrel → Es el error estándar deseado. Define la precisión deseada del estudio.

Deff → Es el efecto del diseño descrito en la Nota 1 de esta sección.

La formulación de los estimadores, los estimadores de varianza y su estructura no se escriben en esta propuesta pero se pueden consultar en textos de muestreo reconocidos como “Model Assisted Survey Sampling” de Carl-Eric Särndal y otros. El Centro Nacional de Consultoría se compromete a entregar la formulación utilizada si expresamente el contratante lo solicita.

Cada una de las unidades estadísticas tiene su factor de expansión. El factor de expansión se construye de acuerdo al diseño de muestreo.

El factor de expansión de hogares (Fh) se obtiene como el producto del factor de manzana dentro de la comuna (etapa I) y del factor de hogar dentro de la manzana (etapa II). El factor de manzana dentro de la comuna es el cociente entre el número de manzanas en el universo y el número de manzanas en la muestra; el factor de hogares es el cociente entre el número total de hogares en la manzana seleccionada y el número de hogares tomados en la muestra de dicha manzana. Adicionalmente, el factor de hogar tiene un componente de ajuste debido a la estructura de hogares por nivel socioeconómico.

El factor de expansión de personas (Fp) parte del factor de hogar y se ajusta con dos componentes básicos: (1) género y (2) grupos etáreos; el objetivo de estos factores de ajuste es garantizar la estructura poblacional por hombres y mujeres y por los siguientes rangos en las edades de las personas: de 0 – 9 años, de 10 a 17 años, de 18 a 24 años, de 25 a 55 años, 56 y más años.

En resumen y de acuerdo a estimativos preliminares, se encuestará a 8.600 hogares; 8.350 de ellos en los cerca de 350 barrios que tiene Cali y 250 hogares en el área rural

(visitando todas las veredas). Las estimaciones resultantes a Total Cali tendrán un error de muestreo no superior a 1.2% y las desagregaciones por comuna, nivel socioeconómico ó población afrodescendiente tendrán asociado un error de muestreo no superior a 7.0%; estos cálculos se realizaron asumiendo una confiabilidad del 95%.

4. Aspectos operativos del diseño de muestreo

Prueba piloto. Se realizará una prueba piloto en 20 hogares en Cali con el fin de detectar posibles inconvenientes en el formulario y de estimar la duración de la aplicación de la encuesta.

No respuesta. La no respuesta es una situación que generalmente afecta a las investigaciones realizadas por muestreo. En este estudio puede entenderse como no respuesta (1) el no diligenciamiento del formulario porque no se encontró a la persona seleccionada ó porque esta no accedió a dar información y el (2) diligenciamiento incompleto del formulario.

En el primer caso se debe buscar un reemplazo, siguiendo las instrucciones del departamento de Estadística de tal manera que el diseño continúe siendo probabilístico; en el segundo caso, es necesario contar con un 2% de formularios completamente diligenciados para reemplazar los incompletos. Si fuera necesario, en los formularios incompletos con poca información faltante, es posible realizar un proceso de imputación de datos sin que el número de formularios con imputación sobrepase el 3% del total de formularios.

Sin embargo, el mejor camino es adoptar mecanismos de control que minimicen la no respuesta. Las siguientes actividades contribuyen a minimizar la no respuesta:

Capacitación	Capacitar debidamente el personal de Campo, brindándole las herramientas que le permitan persuadir a los respondientes para lograr su colaboración. Asegurar que la gente de campo comprende los manuales e instructivos elaborados y que su consulta contribuye a resolver adecuadamente sus dudas. En el caso en que sea inevitable la no respuesta, el personal de campo estará entrenado para tomar debidamente los reemplazos necesarios.
Revisitas	En cada caso de no respuesta en el primer intento, se realizará mínimo una revisita para buscar la información requerida.

Sensibilización y comunicación	Las actividades de sensibilización a la comunidad que realizará la alcaldía de Cali, contribuirá a reducir la tasa de no respuesta.
--------------------------------	---

Dependiendo del nivel de no respuesta final, se analizará la alternativa de implementar los factores de ajuste en las estimaciones finales, debidos a la no repuesta.

Finalmente, se tendrán en cuenta las recomendaciones metodológicas para el operativo de campo contenidas en el capítulo de “Obligaciones de la firma contratada” del documento “Concurso de méritos abierto CMA MT No. 001 de 2012” elaborado por el Ministerio de Trabajo.