

Departamento Administrativo Nacional de Estadística



Dirección de Regulación, Planeación,
Estandarización y Normalización
-DIRPEN-

**Metodología Diseño Estadístico
Encuesta sobre Ambiente y Desempeño
Institucional Nacional
-EDI-**

Junio 2010

	METODOLOGÍA DISEÑO ESTADÍSTICO ENCUESTA SOBRE AMBIENTE Y DESEMPEÑO INSTITUCIONAL NACIONAL -EDI-	CÓDIGO: DM-EDI-MET-01 VERSIÓN : 04 PÁGINA 1 FECHA: 23-06-10
ELABORÓ: DISEÑOS MUESTRALES	REVISÓ: COORDINACIÓN ESTUDIOS ESTADÍSTICOS	APROBÓ :DIRECTOR DIRPEN

CONTENIDO

I. MARCO ESTADÍSTICO.....	2
1.1. POBLACIÓN OBJETIVO DEL DISEÑO MUESTRAL.....	2
1.2. VARIABLES DE INTERÉS.....	2
1.3. PARÁMETROS A ESTIMAR.....	2
1.4. COBERTURA	3
1.5. NIVELES DE DESAGREGACIÓN	3
1.6. UNIDADES ESTADÍSTICAS.....	3
1.7. PERIODO DE REFERENCIA.....	3
1.8. MARCO MUESTRAL	3
1.9. DISEÑO MUESTRAL.....	3
1.9.1 TAMAÑO DE LA MUESTRA	4
1.9.2 DISEÑO Y SELECCIÓN DE LA MUESTRA.....	4
1.9.3 PROCEDIMIENTO DE ESTIMACIÓN.....	6
1.9.4 CÁLCULO DE LA PRECISIÓN DE LOS RESULTADOS	7

I. MARCO ESTADISTICO

1.1. POBLACIÓN OBJETIVO DEL DISEÑO MUESTRAL

Son los funcionarios públicos y empleados, con una antigüedad superior a seis meses, que laboran en la ciudad de Bogotá en las entidades del poder ejecutivo, legislativo, judicial y organismos de control, corporaciones autónomas e institutos de investigación científica, que se cubren en la encuesta.

1.2. VARIABLES DE INTERÉS

- De estudio: Percepción del funcionario frente al ambiente y desempeño institucional en términos de credibilidad en las reglas, en las políticas y frente a los recursos, gestión por resultados, rendición de cuentas y bienestar laboral.
- De clasificación por: Área funcional de las entidades, entidades, nivel de cargo de los funcionarios, edad, sexo, tiempo de servicio.

1.3. PARÁMETROS A ESTIMAR

En los informes publicados, se proporcionan dos tipos de indicadores:

El primer tipo son dos estadísticos que evalúan las respuestas a las preguntas individuales. Uno de los estadísticos consiste en la distribución de frecuencias relativas porcentuales de las respuestas y el otro es un promedio aritmético basado en puntajes de favorabilidad asignados a cada respuesta posible.

El segundo tipo de indicador intenta evaluar las respuestas de un conjunto de preguntas que tienen en común un tema específico (por ejemplo el ambiente institucional). Este tipo de indicadores sintéticos son prácticamente los mismos que los de preguntas individuales. En el caso de la distribución de la frecuencia relativa porcentual, la diferencia radica en la clasificación que se le da a las posibles respuestas. Pues cada respuesta, dependiendo del contexto de la pregunta se le clasifica en cuatro niveles de favorabilidad: Totalmente favorable, favorable, desfavorable, totalmente desfavorable.

En el caso del promedio se construye el promedio de promedios de preguntas individuales que es igual al promedio aritmético de los puntajes de favorabilidad para todas las respuestas de todas las preguntas del tema en específico.

Los parámetros estimados vienen acompañados del coeficiente de variación estimado (cve%) y el intervalo de confianza de la estimación (ic (+-)).

1.4. COBERTURA

La encuesta está dirigida a funcionarios públicos y empleados, con una antigüedad superior a seis meses, de las entidades del poder ejecutivo, legislativo y judicial, organismos de control, ubicados en Bogotá. Se incluye además ISA, ubicada en Medellín, las corporaciones autónomas regionales y las instituciones de investigación científica y ambiental ubicados en el nivel regional.

1.5. NIVELES DE DESAGREGACIÓN

El diseño muestral se realizó para dar estimaciones con niveles de confianza y precisión útiles. Coeficientes de variación menores del 10% para totales y razones estimados a nivel de Entidad, cualquier otra desagregación está sujeta a que las estimaciones no sean tan precisas y su uso dependerá del coeficiente de variación estimado que presenten.

1.6. UNIDADES ESTADÍSTICAS

- Unidad de muestreo: Son los funcionarios de las entidades públicas que tienen más de seis meses de servicio a la entidad y laboran en la sede principal de cada una de las entidades.
- Unidad de observación: Son los mismos funcionarios.
- Unidad de análisis¹: Las entidades públicas.

1.7. PERIODO DE REFERENCIA

El año actual al momento en que se realice la entrevista, de esta forma se ha aplicado para los años 2008 2009 y 2010.

1.8. MARCO MUESTRAL

El marco muestral es el dispositivo que permite identificar y ubicar a cada uno de los elementos de la población objetivo, en este caso a cada uno de los funcionarios de las entidades públicas bajo estudio.

Para construirlo se recurrió a la colaboración de **las Oficinas de recursos Humanos**: de las entidades con quienes se obtuvo la lista de funcionarios, de cada una de ellas.

1.9. DISEÑO MUESTRAL

¹ Ente correspondiente a los diferentes niveles de agregación con el cual se presentan los

El diseño de la muestra es Probabilístico Estratificado de Elementos.

- Probabilístico: Las unidades de muestreo tienen probabilidad de selección conocida y superior a cero de ser seleccionadas.

- Estratificado: El universo se particiona en estratos de acuerdo a los siguientes criterios:

- El primero es la entidad, dado que se requiere entregar información para cada una de las entidades, es necesario garantizar que se selecciona la muestra necesaria para entregar una estimación confiable por lo cual cada una de las entidades se manejó como un estrato.

- EL segundo criterio de estratificación es el nivel de cargo del funcionario, dado que con esto se garantiza que el punto de vista de cada nivel de cargo haga parte del punto de vista de toda la entidad y además se hagan estimaciones confiables a nivel total funcionarios de esta clasificación. Los funcionarios se clasificaron en los siguientes estratos de nivel de cargo.

Nivel 1 Directivos, asesores y ejecutivos

Nivel 2 Profesionales y técnicos

Nivel 3 Operarios y administrativos

1.9.1 TAMAÑO DE LA MUESTRA

Por los resultados obtenidos en la encuesta de los años pasados, se establece que en las entidades donde la cantidad de servidores es inferior a 110, no se hace selección de muestra, sino que se entrevista a todos los servidores de la entidad; por tanto, en estos casos particulares, el valor de las estadísticas calculadas tendrán coeficiente de variación cero ya que no hay error muestral por ser censo.

1.9.2 DISEÑO Y SELECCIÓN DE LA MUESTRA

Para la selección de las unidades de selección probabilística se utilizó muestreo aleatorio simple (MAS) dentro de cada estrato. Para esto se utilizó el método coordinado negativo que consiste en realizar N ensayos con una distribución de probabilidad uniforme (0,1), asignar estos números a cada uno de los elementos del universo, ordenar los elementos respecto a los valores aleatorios y considerar como muestra los elementos correspondientes a los n (tamaño de muestra dentro de cada estrato) valores aleatorios más pequeños.

1.9.3 PROCEDIMIENTO DE ESTIMACIÓN

Los principales parámetros a estimar tienen forma de total y de razón. Sus estimaciones se presentan, según la necesidad de desagregación en dominios o estratos. Ambos son particiones de la población y difieren en la forma en que se debe estimar el parámetro.

Como la base de la estimación se fundamenta en obtener totales por estrato, entonces para el total de una variable y en un estrato h el estimador del parámetro es:

$$\hat{t}_{hy} = \frac{N_h}{n_h} \sum_{s_h} y_k$$

Donde

N_h : Total elementos del universo en el estrato h .

n_h : Total de elementos de la muestra en el estrato.

s_h : Define el conjunto de elementos de la muestra en el estrato h .

y_k : Valor de la variable para el encuestado k .

$\frac{N_h}{n_h}$: Constituye el factor de expansión en el estrato h .

Para la razón entre el total de una variable y y el total de una z en un estrato h el estimador del parámetro es.

$$\hat{R}_h = \frac{\hat{t}_{hy}}{\hat{t}_{hz}}$$

Si la estimación se hace por dominios, entonces el único cambio necesario es el reemplazo de los valores de la variable por cero donde los valores no pertenezcan al dominio.

En la práctica siempre existe la posibilidad de perder información de los encuestados. Esta falta de información genera estimaciones sesgadas de los parámetros. Por lo tanto, dependiendo del tipo de pérdida, se puede corregir el sesgo con un factor de ajuste que se construye por estrato.

Existen dos tipos de pérdida de información en la EDI:

- *Perdida de información causada por los elementos fuera del universo:* Son todos aquellos elementos seleccionados en la muestra que no pertenecen al universo de estudio; entre éstos se cuentan Servidores públicos que ya no pertenecen a la entidad por causa de defunción, jubilación o retiro.

- *Perdida de información causada por la no respuesta:* corresponde a los servidores públicos que, no obstante perteneciendo al universo de estudio, no diligenciaron la encuesta. Dentro de este grupo se encuentran los rechazos.

Tomando en cuenta estos dos tipos de pérdida de información se construye el siguiente factor de ajuste.

$$F_{ajust(h)} = \frac{n_h - n_{fuera\ universo(h)}}{n_h - n_{no\ respuesta(h)} - n_{fuera\ universo(h)}}$$

donde

$n_{fuera\ universo(h)}$: Total servidores públicos seleccionados en el estrato h que NO pertenecen al universo de estudio.

$n_{no\ respuesta(h)}$: Total servidores públicos seleccionados que, aunque pertenecen al universo de estudio en el estrato h, NO responden el cuestionario. En el momento de hacer la estimación se debe tomar el factor de ajuste del estrato y multiplicarlo por el total estimado en ese mismo estrato.

1.9.4 CÁLCULO DE LA PRECISIÓN DE LOS RESULTADOS

El nivel de precisión se fundamenta en la varianza estimada del estimador del parámetro. Para el caso del ESTMAS, el estimador de la varianza del total de una variable y en un estrato h se calcula como:

$$\widehat{VAR}(\hat{t}_{hy}) = \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_{hy}^2$$

Donde

S_{hy}^2 = a la varianza muestral de la variable y en el estrato h .

Por otro lado el estimador de la varianza para el estimador de la razón entre el total de una variable y y el total de una variable z en un estrato h es:

$$\widehat{VAR}(\hat{R}_{hy}) = \widehat{VAR}\left(\frac{\hat{t}_{hy}}{\hat{t}_{hz}}\right) = \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_{hu}^2$$

$$u_{hk} = \frac{1}{\hat{t}_{hz}} (y_{hk} - \hat{R}z_{hk})$$

Donde S_{hu}^2 = a la varianza muestral de la variable u en el estrato h .

Después de tener la varianza estimada del total o de la razón, se puede construir el coeficiente de variación estimado del parámetro como medida de precisión, la idea es que entre menor sea el coeficiente de variación estimado *cve*, menor incertidumbre habrá sobre estimación, su fórmula es:

$$cve = \frac{\sqrt{\hat{V}(\hat{\theta})}}{(\hat{\theta})} * 100$$

Donde $\hat{\theta} = \hat{t}, \hat{R}$ Corresponde al valor estimado del parámetro de interés.
También se puede definir un intervalo de confianza del 95%:

$$IC(\hat{\theta}) = \hat{\theta} \pm 1.96 \sqrt{\hat{V}(\hat{\theta})}$$

Cuando se quiere calcular la varianza estimada del estimador del total para la unión de estratos, se suma las diferentes varianzas estimadas de cada estrato. Lo mismo se hace con estimador de la razón, pero en este caso se debe tomar un u_k construido con la suma de totales estimados de los estratos.

Interpretación de la precisión

Uno de los principales criterios para determinar la calidad de la estimación de un parámetro es la variabilidad que tienen los posibles resultados de dicha estimación. Esta variabilidad se conoce como varianza del estimador, la cual depende de muchos factores, como el diseño muestral, el tamaño de la muestra, el parámetro que se desea estimar, los niveles de desagregación, entre otros.

La varianza se calcula básicamente como la suma del cuadrado de las distancias entre los diferentes valores de una variable y su valor promedio, en este caso, los diferentes valores corresponden a las posibles estimaciones, las cuales, a su vez, provienen de las posibles muestras. Posteriormente, se toma el cuadrado de las distancia para evitar que éstas se anulen entre sí y se disfrace la verdadera dispersión de los datos; por esta razón, la varianza proporciona la variabilidad en unidades al cuadrado, es decir, personas al cuadrado, ganado al cuadrado o hectáreas cuadradas, lo que no permite una comprensión fácil de esta magnitud.

La raíz cuadrada de esta varianza es la que se denomina desviación estándar de la distribución o error estándar. Esta medida de dispersión tiene la ventaja que la unidad de medida de dispersión corresponde a la unidad de la variable de interés, se establece en términos de personas, hectáreas o pesos, aunque queda la dificultad de saber si una desviación es grande o pequeña; así, por ejemplo, una variabilidad de un millón de pesos puede ser muy grande si se habla del promedio de ingresos de los empleados, pero es

absolutamente pequeño si se determina sobre el total del volumen de ventas en la industria del país.

El coeficiente de variación estimado, más conocido como error de muestreo, corresponde al valor que indica el grado de precisión con el cual se está reportando un resultado de las estimaciones de los parámetros definidos con anterioridad. Es decir, se trata de la magnitud de la incertidumbre de una estimación. Se define como la variación porcentual del error estándar a la estimación central, es decir, se trata del cociente entre el error estándar del estimador y el estimador multiplicado por 100, así:

$$CV = \frac{\sqrt{V(\hat{\theta})}}{E(\hat{\theta})} * 100$$

Y el coeficiente de variación estimado está dado por:

$$cve = \frac{\sqrt{\hat{V}(\hat{\theta})}}{(\hat{\theta})} * 100$$

Aunque la varianza, el error estándar y el coeficiente de variación miden la magnitud de la variabilidad de la distribución muestral del estimador, es decir, lo que comúnmente se denomina error de muestreo, el coeficiente de variación tiene la ventaja de proporcionar esta medida en términos porcentuales, por ello se constituye en una medida común para estimaciones.

Se suele considerar que el resultado de una estimación es bueno si su coeficiente de variación es menor de 5%; aceptablemente práctico, entre 5% y 10%; de baja precisión si está en el rango mayor de 10% y menor de 15%; y no útil si es mayor a 15%.

Para entender mejor el significado y los diferentes valores que toman los coeficientes de variación en los cuadros presentados, se debe tener en cuenta que el diseño de la muestra se realizó para obtener estimaciones con alta precisión a nivel nacional por entidad. Las estimaciones para otros niveles de desagregación (como sector) están sujetas a que su precisión no necesariamente sea buena y por tanto el dato no sea confiable.

Es por esta razón que en algunos cuadros aparece, por ejemplo, el total de una variable a nivel nacional y por alguna categoría de análisis con coeficientes de variación pequeños, mientras que para otras categorías de la misma variable los cve son muy altos, en ocasiones del 30 % e incluso mayores del 100 %. En estos casos, el DANE publica la cifra aunque no sea confiable, básicamente para que en los cuadros de salida la información de los totales se observe consistente y porque en muchos casos, el usuario por operaciones aritméticas simples puede deducir el valor correspondiente a esa estimación.



Sin embargo, es muy importante que los usuarios de la información sean conscientes del bajo nivel de precisión que tienen estas estimaciones.

Esta situación puede darse por varias causas. Por ejemplo, cuando el fenómeno estudiado ocurre con gran frecuencia en algunas de las categorías de la variable de clasificación, por lo cual la estimación para dichas categorías es de alta calidad; pero puede ocurrir que para otras categorías en los que el fenómeno no es frecuente la estimación no es buena pues el tamaño de muestra no es suficiente. También puede obedecer al hecho de que en algunas de estas categorías el fenómeno es muy variable mientras en otras es más uniforme, lo que genera menor varianza en las estimaciones.

Todos los resultados que se producen se presentan en cuadros de salida, donde cada estimación tiene su respectivo cve o error muestral. En general, a medida que se incrementa la desagregación de las estimaciones, según las variables de clasificación, el error muestral se incrementa, básicamente porque el tamaño de muestra que incide sobre estos grupos es menor.



BIBLIOGRAFÍA

Bautista S., Leonardo, Diseños de Muestreo Estadístico, Universidad Nacional de Colombia , Departamento de Matemáticas y Estadística, (1998)

Särndal, Carl Erik, Swenson, B. ,Wretman, J. Model Assisted Survey Sampling, Springer-Verlang, New York (1992)