

Departamento Administrativo Nacional de Estadística



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

**Especificaciones de Estimación
Encuesta de Desempeño Institucional
-EDI-**

Marzo 2007

	<p align="center">ESPECIFICACIONES DE ESTIMACIÓN ENCUESTA DE DESEMPEÑO INSTITUCIONAL -EDI-</p>	<p>CÓDIGO: DM-EDI-EES-01 VERSIÓN : 02 PÁGINA 1 FECHA: 23/03/07</p>
<p>ELABORÓ: DISEÑOS MUESTRALES</p>	<p>REVISÓ: COORDINADOR DE PLANIFICACIÓN Y REGULACIÓN</p>	<p>APROBÓ : DIRECTOR DIRPEN</p>

TALBA DE CONTENIDO

3.9.3. METODOLOGÍA DE ESTIMACIÓN Y VARIANZA	5
<u>3.9.3.1. FACTOR DE EXPANSIÓN</u>	<u>6</u>
<u>3.9.3.2. ESTIMACIÓN DE DOMINIOS DE ESTUDIO.....</u>	<u>7</u>
<u>3.9.3.3. AJUSTE DEL FACTOR DE EXPANSIÓN FINAL</u>	<u>11</u>
<u>3.10. COEFICIENTE DE VARIACIÓN.....</u>	<u>12</u>

1. PARÁMETROS A ESTIMAR

En los informes que se publican para cada entidad, se proporcionan dos tipos de indicadores:

El primero que consiste en el promedio aritmético para cada una de las preguntas del cuestionario, calculado sobre las respuestas de los funcionarios entrevistados en cada entidad,

A partir del año 2005 se introdujo otro tipo de indicador que se calcula para dos temas el de ambiente institucional y el de desempeño institucional, este indicador se construye como un promedio a partir del indicador básico, es decir es un promedio aritmético de de promedios aritméticos,

Para comenzar, la información básica con que contábamos contenía, para cada entidad, el promedio de cada pregunta formulada. Así, para el 2004 buscando que los resultados que se obtuvieron pudieran ser comparables con los de la encuesta realizada en convenio con transparencia por Colombia en el 2003 (ver anexo 1. Selección y agrupación de las preguntas EDI 2003 -2004), se seleccionaron únicamente algunas preguntas y se reagruparon en el esquema de ambiente y desempeño institucional que ahora se utiliza para la elaboración de los informes finales. Para el 2005 la base contenía la media de la pregunta P9A hasta la media de la pregunta P29H.

A continuación hay una explicación específica de la forma en que se obtienen los resultados, a partir de una primera aproximación a estos, se debe analizar la base de datos obtenida de la encuesta que contiene promedios por entidad y por pregunta. Ahora bien, en la medida que todos los capítulos constituyen un componente aparte, se calcula el promedio de todas las sentencias que formaban parte de cada uno, en el siguiente cuadro se detalla el número de sentencias por cada componente para el 2005.

COMPONENTE	NÚMERO DE SENTENCIAS. 2005	DE EDI
CREDIBILIDAD EN LAS REGLAS (CREREG)	17	
CREDIBILIDAD EN LAS POLÍTICAS (CREPOL)	27	
CREDIBILIDAD FRENTE LOS RECURSOS (CREREC)	7	
GESTIÓN POR RESULTADOS (GESTRES) ¹	11 (9)	
RENDICIÓN DE CUENTAS (REDCUENT)	40	
BIENESTAR LABORAL (BIENLAB) ²	16 (14)	

¹ Se eliminaron las sentencias 20c y 20d, las cuáles podían causar un error en el cálculo de los promedios, debido a que el planteamiento generaba una inadecuada interpretación por parte de los funcionarios.



**ESPECIFICACIONES DE ESTIMACIÓN
ENCUESTA DE DESEMPEÑO INSTITUCIONAL
-EDI-**

CÓDIGO: DM-EDI-EES-01
VERSIÓN : 02
PÁGINA 3
FECHA: 23/03/07

TOTAL	118
-------	-----

Posteriormente, para cada entidad se calcularon los promedios para cada componente³, de la siguiente manera⁴:

CREREG

$$\sum_1^{17} Xi / 17$$

CREPOL

$$\sum_1^{27} Xi / 27$$

CREREC

$$\sum_1^7 Xi / 7$$

GESTREST

$$\sum_1^9 Xi / 9$$

REDCUENT

$$\sum_1^{40} Xi / 40$$

BIENLAB

$$\sum_1^{14} Xi / 14$$

Así, en Ambiente Institucional se obtuvieron tres promedios: Credibilidad en las Políticas, Credibilidad en la Reglas y Credibilidad frente a los Recursos. En el Caso de Desempeño Institucional se obtuvieron los promedios para Gestión por Resultados, Rendición de Cuentas y Bienestar Laboral. El mismo procedimiento se sigue para el cálculo de los promedios del sector y del Total Nacional.

Por último se obtuvo el puntaje para Ambiente y Desempeño Institucional, que resultó de promediar el resultado obtenido en cada uno de sus componentes. Veamos.

AMBIENTE INSTITUCIONAL

DESEMPEÑO INSTITUCIONAL

² Se eliminaron las sentencias 29d y 29f, debido a que inducían un error en la interpretación por parte de los funcionarios. Se considera que las cuáles podían causar un error en el cálculo de los promedios, debido a que el planteamiento generaba una inadecuada interpretación por parte de los funcionarios.

³ Previamente aquellas preguntas, cuyo valor ideal era 1, se invirtió su escala, para calcular los promedios sobre una escala donde el valor ideal es 5. Este ajuste se llevó a cabo para Credibilidad en las Reglas y Rendición de Cuentas.

⁴ Para el cálculo del promedio, con el fin de unificar el valor ideal en los distintos componentes organizacionales se invirtió el valor de las sentencias 9b, 9c, 9d, 11a, 11b, 11c, 11e, 13 a, 13b, 14 a, 14b, 20g, 20h, 24a-24e, 25a-25 h.26a-26i, 27a-27hy 28f.

$$\left(\sum_1^{17} X_i / 17 + \sum_1^{27} X_i / 27 + \sum_1^7 X_i / 7 \right) / 3 \quad \left(\sum_1^9 X_i / 9 + \sum_1^{40} X_i / 40 + \sum_1^{14} X_i / 14 \right) / 3$$

Gráficamente, el resultado obtenido para cada entidad se presentaría como dos tríadas; una para el Ambiente Institucional y otra para el Desempeño Institucional, cada una se compara con la tríada ideal, la tríada para el sector y la tríada del Total Nacional. Los gráficos 1 y 2 presentan un ejemplo del resultado general de una entidad hipotética.

Este mismo procedimiento se siguió para obtener las tríadas de ambiente y desempeño discriminadas por nivel y antigüedad. A continuación se muestra gráficamente este análisis:

GRÁFICO 1. AMBIENTE INSTITUCIONAL (EJEMPLO)

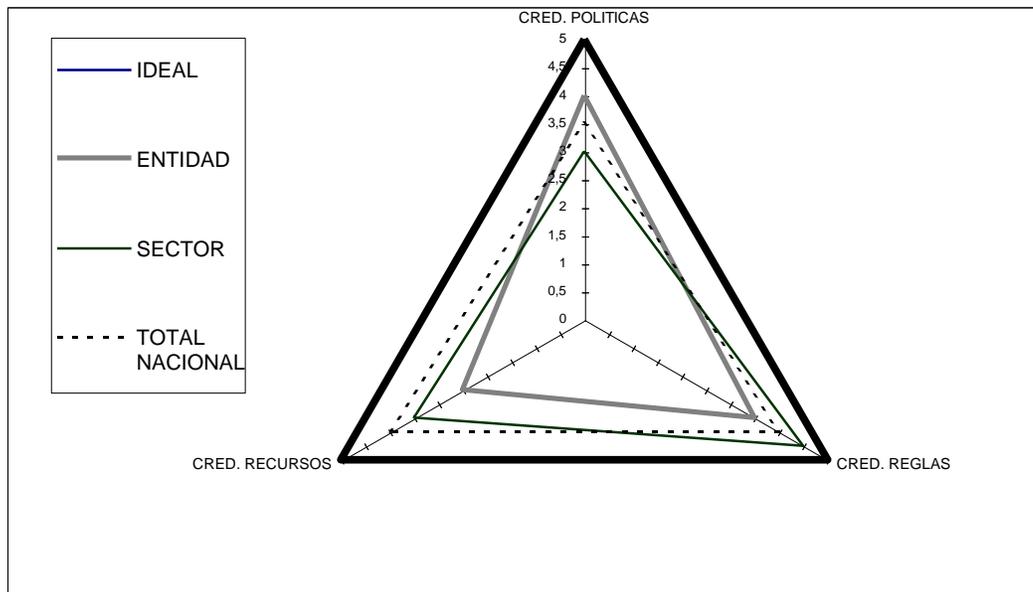
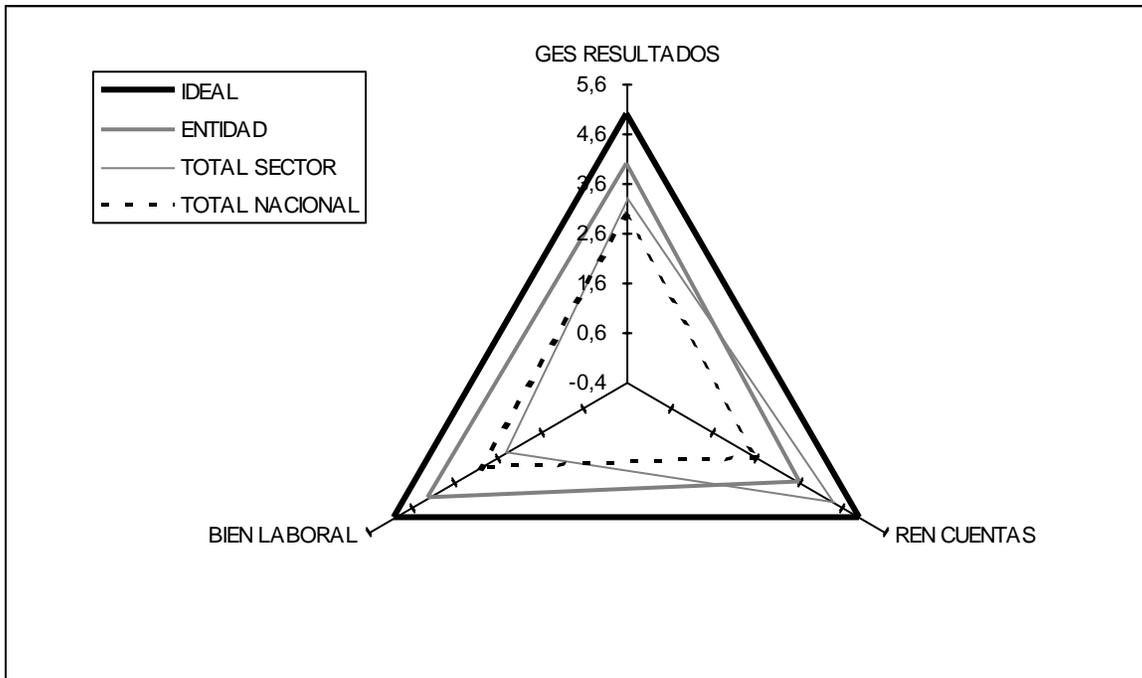


GRÁFICO 2. DESEMPEÑO INSTITUCIONAL (EJEMPLO)



Para contribuir a la interpretación de los resultados de Ambiente y Desempeño Institucional, en la guía de interpretación de resultados, que se adjunta a los informes de cada entidad, se mencionan las pautas de comparación entre entidades, niveles y antigüedad.

Finalmente para cada uno de los principios que permiten captar la estructura de gobernanza de una entidad, se escogieron sentencias de los distintos capítulos de la encuesta con el fin de consolidar información que permita visualizar los aspectos positivos y negativos de la entidad durante los 12 meses previos a la aplicación del formulario de la EDI.

El cálculo de coeficiente de variación, límites inferior y superior se calcula para el promedio aritmético simple por pregunta (estos últimos se encuentran para el 2004 al interior de los informes de resultados de cada entidad y para el 2005 en los archivos anexos que tienen por título errores de muestreo).

2. METODOLOGÍA DE ESTIMACIÓN Y VARIANZA

Como se mencionó anteriormente los principales parámetros a estimar en la encuesta de desempeño institucional son promedios aritméticos, con el principio básico que cada

elemento de la muestra da una calificación de la entidad y luego se estima el promedio para el total de funcionarios de la entidad.

A continuación se presenta la metodología de estimación y varianza para los promedios aritméticos, que se entregan para cada pregunta a nivel de entidad.

Como se trata de un diseño E.S.T. M.A.S. (estratificado donde al interior de cada estrato el mecanismo de selección fue M.A.S) el factor de expansión esta dado por:

2.1. FACTOR DE EXPANSIÓN

El factor de expansión es un valor que multiplica a cada uno de los elementos seleccionados en la muestra con el fin de obtener la estimación del parámetro en el universo, este factor se depende y se construye a partir del método de selección utilizado.

Para el diseño ESTMAS el factor de expansión, se construye como el inverso de la probabilidad de inclusión y esta dado por:

$$F_h = N_h / n_h$$

N_h = Cantidad de elementos del universo en el estrato h

n_h = Cantidad de elementos de la muestra en el estrato h

Dado que existen novedades en el proceso de recolección, causando perdida de muestra que afecta directamente la estimación es necesario realizar ajuste por algunas razones a este factor de expansión.

Existen tres grandes tipos de novedad.

- Primera Los fuera de universo: En este grupo se incluyen todos aquellos elementos seleccionados en la muestra que no pertenecen al universo de estudio, entre ellos se cuentan, para esta encuesta en particular, Los jubilados, retirados, las personas que laboran fuera de la ciudad, las personas que llevan menos de un año laborando en la institución, etc,
- Segundo la no respuesta: (deuda) Corresponde a las personas que perteneciendo al universo de estudio no diligenciaron la encuesta, dentro de este grupo se encuentran los que estaban en vacaciones, licencias médicas, comisión de trabajo, comisión , incapacidad médica ,secuestrado o privado de la libertad, reuniones en otra entidad, descanso compensatorio, licencia no remunerada, permiso o licencia remunerada, analfabetas y quienes no quisieron diligenciar el cuestionario.

- Tercero los adicionales: Corresponde a las personas que perteneciendo al universo de estudio no se encontraban relacionados en el marco muestral, pero que al realizar la encuesta en campo se detectó el subregistro y se les aplicó la entrevista.

El factor de ajuste esta dado por:

$$F_{ajust} = \frac{n_h + n_{hadi} - n_{hfu}}{n_h + n_{hadi} - n_{hfu} - n_{deuda}}$$

donde

n_h = Cantidad de personas seleccionadas en el estrato h .

n_{hdeuda} = Cantidad de personas en el estrato h adicionadas, corresponde a aquellas que perteneciendo al universo de estudio no se encontraban en el marco de muestreo.

n_{hdeuda} = Cantidad de personas seleccionadas en el estrato h que no respondieron el cuestionario, perteneciendo al universo de estudio

n_{hfu} = Cantidad de personas seleccionadas en el estrato h que no pertenecían a l universo de estudio

Finalmente el factor final de expansión esta dado por

$$F_{final} = F_h * F_{ajust}$$

2.2. ESTIMACIÓN DE DOMINIOS DE ESTUDIO

Un Dominio de Estudio es una subpoblación para la cual se requieren estimaciones puntuales separadas con buena precisión y con intervalos de confianza útiles. Si bien la muestra esta diseñada para entregar indicador a nivel entidad y muy posiblemente no se puedan realizar estimaciones más desagregadas por entidad, si es posible realizar estimaciones que agreguen un grupo de entidades, por ejemplo los dominios de estudio podrían ser:

- Percepción total entidades públicas por nivel jerárquico
- Percepción entidades poder ejecutivo por rango de edad
- Percepción por rama de la entidad (ejecutivo, legislativo, judicial)
- Percepción por rama de la entidad y antigüedad en la institución, etc.

En estos casos el algoritmo de estimación es:

Sea la variable Z_{dk} definida como

$$Z_{dk} = \begin{cases} 1 & \text{si } k \in U_d \\ 0 & \text{si } k \notin U_d \end{cases}$$

donde

k = Funcionario público

Ud = Dominio d

Luego

$$\sum_U z_{dk} = N_d$$

Nd = Cantidad de elementos en el universo que pertenecen al dominio d, la cual es desconocida.

Ahora sea la variable

$$y_{dk} = x_{dk} * z_{dk}$$

x_k = Es la variable cuantitativa de interés y observada en la muestra

Bajo el diseño de muestreo estratificado,

El total de un dominio es:

t_d = total de la variable x en el dominio d

Parámetro

$$t_d = \sum_U x_{dk}$$

Estimador

$$\begin{aligned} \hat{t}_d &= \sum_{h=1}^H N_h \bar{y}_{dh} \\ &= \sum_{h=1}^H \sum_{m_h} F_h y_{hk} \end{aligned}$$

Con

$$\bar{y}_{d_h} = \sum_{m_h} y_{hk} / n_h$$

Nota: Los subíndices de las sumas se refieren a:

U = Universo muestral

m_h = Las unidades seleccionadas en la muestra que pertenecen al estrato h
 y_{hdk} = Al valor de la variable y para el elemento k en el estrato h en el dominio d
 $y_{hd\bar{m}}$ = Al promedio de la variable y en el estrato h para el dominio d
 n_{dh} = Al tamaño de la muestra para el estrato h en el dominio d
 F_h = El factor de expansión para los elementos del estrato h

La varianza del total de un dominio esta dada por

$$V(\hat{t}) = \sum_{h=1}^H N_h^2 \frac{1-f_h}{n_h} S_{y d U_h}^2$$

Donde $f_h = \frac{n_h}{N_h}$ es la fracción de muestreo en el estrato h y

$$S_{y d U_h}^2 = \frac{1}{N_h - 1} \sum_{U_h} (y_{dk} - \bar{y}_{dU_h})^2$$

es la varianza de la variable y para el dominio d en el estrato h

Un estimador insesgado de la varianza es

$$\hat{V}(\hat{t}) = \sum_{h=1}^H N_h^2 \frac{1-f_h}{n_h} \hat{S}_{y d h}^2$$

Donde la varianza muestral o estimada de la variable y para el dominio d en el estrato h es:

$$\hat{S}_{y d h}^2 = \frac{1}{n_h - 1} \sum_{m_h} (y_{dk} - \hat{y}_{d h})^2$$

Estimadores de razón para un dominio:

Un estimador de razón para un dominio es de la forma

$$R = \frac{t_{dy}}{t_{dx}}$$

donde

t_{dy} Es el total de la variable y en el dominio d
 t_{dx} Es el total de la variable x en el dominio d

Y el estimador es

$$\hat{R} = \frac{\hat{t}_{dy}}{\hat{t}_{dx}}$$

t^{\wedge}_{dy} Es el total estimado de la variable y en el dominio d
 t^{\wedge}_{dx} Es el total estimado de la variable x en el dominio d

La varianza del estimador de razón

Dado que los estimadores de razón son funciones no lineales de totales, su estimador es sesgado aunque con un sesgo relativo no muy grande, la varianza aproximada del estimador de razón esta dada por

$$AV(\hat{R}_d) = \sum_{h=1}^H \frac{1(1-f_h)}{\bar{x}_{Ud}^2 n_h} (S_{xUdh}^2 + R_d^2 S_{yUdh}^2 - 2R_d S_{yxUdh})$$

con

$$S_{xUdh}^2 = \frac{1}{n_h} \sum_{U_h} (x_{dk} - \bar{x}_{dU_h})^2$$

la varianza poblacional de la variable x para el dominio d en el estrato h

$$S_{yUdh}^2 = \frac{1}{n_h} \sum_{U_h} (y_{dk} - \bar{y}_{dU_h})^2$$

la varianza poblacional de la variable y para el dominio d en el estrato h

$$S_{xyUdh}^2 = \frac{1}{n_h} \sum_{U_h} (x_{dk} - \bar{x}_{dU_h})(y_{dk} - \bar{y}_{dU_h})$$

la covarianza poblacional de las variable x y para el dominio d en el estrato h

El estimador de la varianza del estimador de razón es:

$$\hat{V}(\hat{R}_d) = \sum_{h=1}^H \frac{1(1-f_h)}{\hat{x}_d^2 n_h} (\hat{S}_{xUdh}^2 + \hat{R}_d^2 \hat{S}_{yUdh}^2 - 2\hat{R}_d \hat{S}_{yxUdh})$$

$$\hat{S}_{x_{dh}}^2 = \frac{1}{n_h - 1} \sum_{m_h} (x_{dk} - \hat{x}_{dh})^2$$

con la varianza poblacional de la variable x para el dominio d en el estrato h

$$\hat{S}_{y_{dh}}^2 = \frac{1}{n_h - 1} \sum_{m_h} (y_{dk} - \hat{y}_{dh})^2$$

la varianza poblacional de la variable y para el dominio d en el estrato h

$$\hat{S}_{xy_{dh}}^2 = \frac{1}{n_h - 1} \sum_{m_h} (x_{dk} - \hat{x}_{dh})(y_{dk} - \hat{y}_{dh})$$

la covarianza poblacional de las variable x y para el dominio d en el estrato h

2.3. AJUSTE DEL FACTOR DE EXPANSIÓN FINAL

Dado que existen novedades en el proceso de recolección, causando perdida de muestra que afecta directamente la estimación es necesario realizar ajuste por algunas razones a este factor de expansión.

Existen tres grandes tipos de novedad, los fuera de universo, la no respuesta y los adicionales

El factor de ajuste esta dado por:

$$F_{ajust} = \frac{n_h + n_{hadi} - n_{hfu}}{n_h + n_{hadi} - n_{hfu} - n_{deuda}}$$

donde

n_h = Cantidad de personas seleccionadas en el estrato h.

n_{hdeuda} = Cantidad de personas en el estrato h adicionadas, corresponde a aquellas que perteneciendo al universo de estudio no se encontraban en el marco de muestreo.

n_{hdeuda} = Cantidad de personas seleccionadas en el estrato h que no respondieron el cuestionario, perteneciendo al universo de estudio

$n_{hfu} =$ Cantidad de personas seleccionadas en el estrato h que no pertenecían a l universo de estudio

Finalmente el factor final de expansión esta dado por

$$F_{final} = F_h * F_{ajust}$$

3. COEFICIENTE DE VARIACIÓN.

Cuando la magnitud de la variabilidad es muy grande los parámetros estimados pierden utilidad, pues el valor verdadero del parámetro, en el universo, puede estar en un intervalo muy amplio lo cual no proporciona información útil. Por lo tanto esta medida de variabilidad es un criterio muy importante para determinar la calidad de una estimación.

Pero el valor de esta magnitud esta directamente relacionada con la unidad de medida de la variable de interés, por ejemplo si se esta estimando el total de los ingresos para una actividad económica dada la unidad de medida en miles de pesos, y como además la varianza está en unidades al cuadrado, se estaría hablando de miles de pesos al cuadrado. 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 de que la unidad de la medida de dispersión corresponde a la unidad de la variable de interés, así por ejemplo, una variabilidad de cien en un millón en la estimación del total es pequeña lo cual indicaría que la estimación tiene muy buena precisión, pero si el parámetro a estimar es el promedio de personas ocupadas por empresas, tener una variabilidad de cinco personas es alta, y aunque es un valor muchísimo menor puede significar que no es tan preciso.

Un criterio más uniforme para determinar la precisión de un estimador es el coeficiente de variación el cual se define como la variación porcentual del error estándar a la estimación central, es decir es el cociente entre el error estándar del estimador y el estimador multiplicado por 100.

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

el coeficiente de variación estimado esta dado por

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



**ESPECIFICACIONES DE ESTIMACIÓN
ENCUESTA DE DESEMPEÑO INSTITUCIONAL
-EDI-**

CÓDIGO: DM-EDI-EES-01
VERSIÓN : 02
PÁGINA 13
FECHA: 23/03/07

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 algunos denominan el error de muestreo, el coeficiente de variación tiene la ventaja de dar esta medida en términos porcentuales, por lo cual es una medida común para estimaciones.

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

Para el mejor entendimiento del significado y los diferentes valores que toman los coeficientes de variación en los cuadros presentados se debe tener en cuenta los siguientes aspectos.

El diseño de la muestra se realizó para obtener estimaciones con alta precisión a nivel entidad. Cualquier otro nivel de desagregación o parámetro estimado esta sujeto a que su precisión no necesariamente sea buena y por tanto el dato no sea publicable.



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)