

República de Panamá
CONTRALORIA GENERAL DE LA REPÚBLICA
Instituto Nacional de Estadística y Censo
Unidad de Muestreo

METODOLOGÍA DEL DISEÑO DE MUESTREO

Encuesta entre Empresas no Financieras 2013

1. El objetivo de la EENF es generar estimaciones anuales, por rama de actividad, que permitan conocer los ingresos de las empresas no financieras con el propósito de determinar su contribución al Producto Interno Bruto (PIB), así como la demanda de mano de obra.

De manera específica, es de interés conocer los ingresos totales de las empresas no financieras en los siguientes sectores de actividad económica.

- Silvicultura, extracción de madera y actividades en servicios conexos
- Explotación de minas y canteras
- Industria Manufacturera
- Suministro de electricidad, gas y agua
- Comercio al por mayor
- Comercio al pormenor
- Hoteles y restaurantes
- Transporte, almacenamiento y comunicaciones
- Actividades inmobiliarias empresariales y de alquiler
- Enseñanza
- Actividades de servicios sociales y salud
- Otras actividades comunitarias, sociales y personal de servicio

2. De acuerdo a las decisiones asumidas para diseñar la encuesta se aplicaron los siguientes criterios técnicos:

- Se seleccionaron con certeza las empresas que manifiestan disponer de inversión extranjera directa (IED).
- Asimismo, también se incluyeron con certeza las empresas que pertenecen a una rama de actividad conformada por 5 o menos unidades de observación.
- Fueron incorporadas con probabilidad 1 por rama de actividad, las empresas que declararon ingresos superiores a un cierto monto definido.

- Al respecto, cabe aclarar que no se asumió un criterio uniforme para conformar los estratos por rama de actividad. El valor del umbral asumido se ubicó en función al monto máximo de ingreso reportado por las empresas en cada rama.
- El resto de las empresas, que no se incluyeron con certeza, formó parte del marco de muestreo.
- De este grupo de empresas se seleccionó la muestra probabilística que sustenta el actual diseño estadístico que se aplica en la EENF.

3. Dominios de estudio

Está constituido por las siguientes provincias: Bocas del Toro, Coclé, Colón, Chiriquí, Darién, Herrera, Los Santos, Panamá, Veraguas.

4. Diseño de la muestra

El procedimiento estadístico que se definió para la selección de la muestra de empresas de la EENF, corresponde a un diseño estratificado aleatorio, a partir de una doble estratificación.

En este tipo de investigaciones se debe garantizar, en la medida de lo posible que la unidad de observación, selección y análisis sea la empresa, que habitualmente se define como el grupo de dos o más establecimientos que llevan a cabo actividades económicas no financieras.

Cuando se corrobora que las empresas están conformadas por dos o más establecimientos la instrucción es recopilar información de todas las unidades que la conforman.

5. Estratificación del marco de muestreo

La primera agrupación se definió utilizando la rama de actividad económica principal de la empresa como criterio, en tanto que la subestratificación se aplicó utilizando como variable de corte el monto del ingreso total anual reportado por las empresas.

Después de obtener los tamaños de muestras por ramas de actividad, se construyen estratos en la rama donde haya necesidad de estratificar de acuerdo a la regla de Dalenius y Hodges.

6. Marco de Muestreo

El marco muestral está basado en los resultados de los Sextos Censos Nacionales Económicos 2012, por ramas de actividad según provincias.

7. Unidades de Muestreo

Las unidades de muestreo son las empresas.

8. Tamaño de la Muestra

Conforme a los requerimientos de la encuesta, el diseño muestral de la EENF permite generar estimaciones por rama de actividad (cuatro dígitos), según provincias.

La variable de diseño utilizada para determinar el tamaño de la muestra corresponde al ingreso total de las empresas el cual se obtuvo de la información recopilada en los Censos Nacionales Económicos 2012, por rama de actividad según provincia.

El número de empresas seleccionadas por estrato (h) se determinó de manera independiente en cada rama de actividad (j) a partir de la expresión (1) que tiene en cuenta el número de empresas en cada estrato N_{jh} , su varianza S_{jh}^2 , el error máximo esperado (e) y la confianza deseada para las estimaciones que se expresa a partir del valor que asume la variable (z) que representa el valor en tablas de una distribución normal que en este caso se fijó en 95%.

$$n_j = \frac{\left(\sum_{h=1}^L N_{jh} S_{jh} \right)^2}{\frac{e^2}{z^2} + \sum_{h=1}^L N_{jh} S_{jh}^2} \quad (1)$$

Observe que el tamaño total de la muestra se genera a partir de estimar el error de muestreo de cada estrato considerando los niveles de precisión y confianza deseados.

Para aplicar la expresión (1) se utilizó la información contenida en el Censo Económico del 2012 y el cálculo de la varianza se llevó a cabo a partir del ingreso total reportado por las empresas que se definió como variable de diseño.

9. Asignación de la muestra por estrato

Para la asignación de la muestra por estrato se utilizó el criterio de asignación óptima de Neyman, teniendo en cuenta que la dispersión (varianza) del ingreso difiere por estrato.

Es decir, de acuerdo a la teoría se sugiere asignar la muestra con este criterio cuando el propósito es mejorar la precisión de los estimadores teniendo en cuenta la variabilidad del indicador de interés al interior de los estratos.

Conforme a lo anterior, la determinación del tamaño de muestra por estrato se asignó utilizando la siguiente expresión.

$$n_{jh} = n_j * \frac{N_{jh} S_{jh}}{\sum_{h=1}^4 N_{jh} S_{jh}} \quad (2)$$

En donde n_{jh} representa el tamaño de muestra del estrato h en la rama de actividad j, en tanto que S_{jh} da cuenta de la desviación estándar del ingreso total utilizado para el cálculo en el estrato h en la rama de actividad j.

De esta manera, se seleccionaron 2,905 empresas divididos de la manera en que se muestra en el cuadro 1: 1,962 empresas se incluyeron con certeza y el resto se eligieron conforme al diseño probabilístico descrito.

**Tabla 1 ENCUESTA ENTRE EMPRESAS NO FINANCIERAS, AÑO 2014
TOTAL DE LA MUESTRA POR ESTRATO
SEGÚN DOMINIO DE ESTUDIO**

DOMINIO DE ESTUDIO	TOTAL	ESTRATO				
		I	II	III	IV	V
TOTAL	2,905	1,962	559	223	113	48
BOCAS DEL TORO	74	68	5	1	0	0
COCLE	157	128	23	6	0	0
COLON	163	156	7	0	0	0
CHIRIQUI	425	284	72	35	34	0
DARIEN	11	11	0	0	0	0
HERRERA	129	128	1	0	0	0
LOS SANTOS	99	96	3	0	0	0
PANAMA	1,669	916	445	181	79	48
VERAGUAS	178	175	3	0	0	0

10. Selección de los establecimientos en muestra en cada estrato

En cada uno de los estratos formados se seleccionó una muestra independiente de tamaño n_{jh} a partir de un procedimiento de muestreo sistemático con arranque aleatorio. Es decir, se asignó la misma probabilidad de selección a las empresas de cada estrato.

De esta manera, el procedimiento de selección que se aplicó fue el siguiente.

Se definió un intervalo de selección $k = \frac{N_{jh}}{n_{jh}}$ el cual se forma dividiendo el número de empresas en el estrato h de la rama j entre tamaño de muestra en el estrato h de la rama j (n_{jh}).

Posteriormente se seleccionó un número aleatorio (x) entre 1 y k y al valor obtenido se le adicionó el valor de k ($x+k$) hasta obtener el total de empresas que deben elegirse por estrato.

11. Cálculo de los factores de expansión

Las empresas se seleccionaron en una etapa aplicando muestreo aleatorio simple. De esta manera, el factor de expansión asociado a cada empresa de la rama j en el estrato h se obtiene al dividir el total de empresas en el estrato h de la rama j (N_{jh}) entre el total de empresas seleccionadas en el estrato h de la rama de actividad j (n_{jh}), de la manera en que se muestra a continuación.

$$f_{jh} = \frac{N_{jh}}{n_{jh}} \quad (3)$$

De esta manera se observa que, al definirse los factores de expansión como el inverso de la probabilidad de selección de las empresas en cada estrato debe satisfacerse la premisa de que su valor no debe ser superior a la unidad.

Esta afirmación es válida sobre todo en el caso de las empresas que fueron seleccionadas con certeza.

12. Estimador de totales

Sea y_{jkh} el valor de la variable de interés en la empresa k del subestrato h de la rama j y w_{jkh} el factor de expansión correspondiente. De esta manera, el estimador muestral del total de la variable y se define como:

$$\hat{Y} = \sum_k w_{jkh} y_{jkh} \quad (4)$$

13. Estimadores para promedios

El estimador de una media poblacional de la variable de interés y se obtiene como el cociente entre el estimador del total y la estimación del tamaño de la población:

$$\hat{\bar{y}} = \frac{\hat{Y}}{\hat{N}} = \frac{\sum w_{jkh} y_{jkh}}{\sum w_{jkh}} \quad (5)$$

14. Precisión de los estimadores

Los errores de muestreo ocurren debido a que las estimaciones se sustentan en información obtenida de una muestra. En este sentido, el diseño aplicado, la variabilidad de los datos y el tamaño de muestra determinan el error de estimación.

Una medida que se utiliza comúnmente para aproximar el error de las estimaciones es la desviación estándar. Este indicador de cuenta del grado de variación en las estimaciones como resultado de la selección de una muestra en particular en lugar de otra del mismo tamaño y diseño.

El verdadero valor del parámetro de interés se ubica, por ejemplo, dentro de un rango de +/- una desviación estándar en el 68 % de las muestras y de +/- dos desviaciones estándar para el 95 % aproximado de todas las muestras.

El coeficiente de variación (*CV*) representa la desviación estándar expresada como un porcentaje del valor del estimador. Se utiliza para indicar el grado de incertidumbre asociado a una estimación. Por ejemplo, si la estimación puntual del Personal Ocupado para una división de actividad es 1000 y el *CV* correspondiente es del 5%, entonces el verdadero valor se encuentra del estimador se ubica entre 950 y 1050, en el 68 % de las veces y entre 90 y 1100 en el 95 % de las veces.

Los errores de muestreo de la EENF_2014 se pueden calcular utilizando, por ejemplo, el módulo de encuestas complejas del SPSS, STATA o SAS.