

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

DIRECCIÓN GENERAL DE ESTADÍSTICAS ECONÓMICAS  
DIRECCIÓN DE MARCOS Y MUESTREO

Nota Metodológica  
sobre Diseño Muestral

Encuesta Nacional de Empresas Constructoras 2012  
(Mensual y Anual)



INSTITUTO NACIONAL  
DE ESTADÍSTICA Y GEOGRAFÍA



## Índice

---

<b>1. Objetivo de la encuesta</b>	<b>1</b>
<b>2. Población objetivo</b>	<b>1</b>
<b>3. Cobertura sectorial</b>	<b>1</b>
<b>4. Dominios de estudio</b>	<b>1</b>
<b>5. Diseño de la muestra</b>	<b>1</b>
5.1 Marco de la encuesta	1
5.2 Unidad de muestreo	1
<b>6. Esquema de muestreo</b>	<b>1</b>
6.1 Formación de la componente principal	1
6.2 Estratificación de la componente principal	1
<b>7. Tamaño de la muestra</b>	<b>2</b>
<b>8. Afijación de la muestra</b>	<b>2</b>
<b>9. Selección de la muestra</b>	<b>3</b>
<b>10. Cálculo de los factores de expansión</b>	<b>3</b>
10.1 Ajuste por no Respuesta	3
<b>11. Estimadores</b>	<b>4</b>
<b>Anexo</b>	<b>5</b>



## **1. Objetivo de la encuesta**

Generar información estadística que permita elaborar una serie de indicadores económicos con el propósito de conocer la evolución y comportamiento de la actividad económica que realizan las empresas constructoras en el país.

## **2. Población objetivo**

La población objeto de estudio son las empresas constructoras dedicadas a la edificación residencial, a la construcción de obras de ingeniería civil y a la realización de trabajos especializados, clasificadas en el sector 23 Construcción según el Sistema de Clasificación Industrial de América del Norte 2007 (SCIAN 2007).

## **3. Cobertura sectorial**

La cobertura de la encuesta permite obtener estimaciones nacionales por subsector de actividad económica según el SCIAN 2007 y por entidad federativa.

## **4. Dominios de estudio**

Cada subsector de actividad económica a nivel nacional y cada entidad federativa son dominios de estudio.

## **5. Diseño de la muestra**

El diseño de la muestra se caracterizó por ser probabilístico y estratificado.

### **5.1. Marco de la encuesta**

El marco de muestreo está formado por las empresas del directorio definitivo de los Censos Económicos 2009, clasificadas en el sector 23 Construcción según el SCIAN 2007, actualizado con base en la información proveniente del Directorio Nacional de Unidades Económicas (DENUE) y de la misma encuesta.

El marco quedó integrado por 22 106 empresas (ver Anexo).

### **5.2. Unidad de muestreo**

La unidad de muestreo fue la empresa constructora.

## **6. Esquema de muestreo**

El esquema de muestreo es probabilístico y estratificado, con selección independiente para cada estrato por dominio de estudio.

### a) Probabilístico

Las empresas tuvieron una probabilidad conocida y distinta de cero de ser seleccionadas.

### b) Estratificado

Las unidades de muestreo con características similares se agruparon para formar estratos.

### **6.1. Formación de la componente principal**

Para cada empresa del marco de muestreo, se obtuvo la primer componente principal, la cual es una combinación lineal de las variables personal ocupado total y valor de la producción.

### **6.2. Estratificación de la componente principal**

Se estableció una clasificación a cinco estratos, aplicando el algoritmo de k medias a la componente principal, excepto para el primer estrato en el que se incluyeron con certeza a las empresas con más de 80 millones de pesos en el valor de la producción o personal ocupado mayor o igual a 250.

Se obtuvo la siguiente distribución:

Cuadro 1

Estrato	Límites	Empresas
<b>Total</b>		<b>22 106</b>
1	6.29490 <= CP	731
2	2.09712 <= CP < 6.29490	725
3	-0.73162 <= CP < 2.09712	1 339
4	-0.20806 <= CP < -0.73162	3 854
5	CP < -0.20806	15 457

## 7. Tamaño de la muestra

Para el cálculo de tamaño de muestra se utilizó como variable de referencia a la componente principal, considerando un nivel de confianza del 95%, error relativo del 2% al 11% según el dominio de estudio y una tasa de no Respuesta esperada del 20%.

La expresión empleada para cada dominio fue:

$$n_e = \frac{\left( \sum_{h=1}^H N_{eh} \sigma_{eh} \right)^2}{\left( \frac{r^2 T_e^2}{z^2} + \sum_{h=1}^H N_{eh} \sigma_{eh}^2 \right) (1 - TNR)}$$

Donde:

- $n_e$  = tamaño de muestra en el e-ésimo dominio.
- $H$  = número de estratos por dominio.
- $N_{eh}$  = total de empresas en el h-ésimo estrato, en el e-ésimo dominio.
- $\sigma_{eh}$  = desviación estándar de la componente principal del h-ésimo estrato, del e-ésimo dominio.
- $r$  = error relativo.
- $T_e$  = total de la componente principal en el e-ésimo dominio.
- $z$  = valor de la abscisa bajo distribución normal estándar para un nivel de confianza prefijada.
- $\sigma_{eh}^2$  = varianza de la componente principal del h-ésimo estrato, del e-ésimo dominio.
- $TNR$  = tasa de no Respuesta máxima esperada.

El tamaño de muestra total fue de 3 594 empresas, considerando a las empresas incluidas con certeza. (Ver anexo)

## 8. Afijación de la muestra

Para cada dominio de estudio, se distribuyó la muestra por estrato considerando la afijación de Neyman tomando como base la ponderación de la componente principal de la siguiente manera:

$$n_{eh} = n_e \frac{N_{eh} \sigma_{eh}}{\sum_{h=1}^H N_{eh} \sigma_{eh}}$$

Donde:

- $n_{eh}$  = tamaño de muestra asignada en el h-ésimo estrato, en el e-ésimo dominio.
- $n_e$  = tamaño de la muestra en el e-ésimo dominio.

- $N_{eh}$  = total de empresas en el marco en el h-ésimo estrato, en el e-ésimo dominio.  
 $\sigma_{eh}$  = desviación estándar de la componente principal del h-ésimo estrato, del e-ésimo dominio.  
 $H$  = total de estratos por dominio.

### 9. Selección de la muestra

Las empresas se seleccionaron de manera aleatoria, sin reemplazo e independiente para cada dominio-estrato.

### 10. Cálculo de los factores de expansión

El factor de expansión fue el inverso de la probabilidad de seleccionar una empresa del h-ésimo estrato, del e-ésimo dominio.

$$F_{eh} = \frac{N_{eh}}{n_{eh}}$$

Donde:

- $F_{eh}$  = factor de expansión del h-ésimo estrato, del e-ésimo dominio.  
 $N_{eh}$  = total de empresas en el marco en el h-ésimo estrato, en el e-ésimo dominio.  
 $n_{eh}$  = total de empresas seleccionadas en la muestra en el h-ésimo estrato, en el e-ésimo dominio.

### 10.1. Ajuste por no Respuesta

El ajuste del factor de expansión por no Respuesta se aplica cuando no se realizan todas las entrevistas esperadas, y se calcula con base en la expresión:

$$f'_{eh} = f_{eh} \frac{n_{eh}}{r'_{eh}}$$

Donde:

- $f'_{eh}$  = factor de expansión ajustado por no Respuesta del h-ésimo estrato, del e-ésimo dominio.  
 $f_{eh}$  = factor de expansión del h-ésimo estrato, del e-ésimo dominio.  
 $n_{eh}$  = total de empresas seleccionadas en la muestra en el h-ésimo estrato, en el e-ésimo dominio.  
 $r'_{eh}$  = total de empresas en la muestra recuperada en el h-ésimo estrato, en el e-ésimo dominio.

### 11. Estimadores

El estimador del total de la característica X en el e-ésimo dominio es:

$$\hat{t}_e = \sum_{h=1}^H \sum_{i=1}^{n_{eh}} f'_{eh} X_{ehi}$$

El estimador de la varianza es:

$$\hat{V}(\hat{t}_e) = \sum_{h=1}^H N_{eh}^2 \left(1 - \frac{n_{eh}}{N_{eh}}\right) \frac{S_{eh}^2}{n_{eh}} \quad \text{donde:} \quad S_{eh}^2 = \frac{1}{n_{eh}-1} \sum_{i=1}^{n_{eh}} [X_{ehi} - \bar{X}_{eh}]^2$$

Error de estimación:  $EE = \sqrt{\hat{V}(\hat{t}_e)}$

Coeficiente de variación: 
$$CV = \frac{EE}{\hat{t}_e}$$

Intervalo de confianza: 
$$\left( \hat{t}_e - Z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{t}_e)}, \hat{t}_e + Z_{\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{t}_e)} \right)$$

Donde:

- $\hat{t}_e$  = total estimado en el e-ésimo dominio.
- $H$  = total de estratos por dominio.
- $n_{eh}$  = tamaño de la muestra en el h-ésimo estrato, en el e-ésimo dominio.
- $f'_{eh}$  = factor de expansión ajustado por no Respuesta del h-ésimo estrato, del e-ésimo dominio.
- $X_{ehi}$  = el valor de la característica de interés en la i-ésima empresa, en el h-ésimo estrato, en el e-ésimo dominio.
- $\hat{V}(\hat{t}_e)$  = varianza de diseño estimada del total estimado del e-ésimo dominio.
- $N_{eh}$  = tamaño de la población en el h-ésimo estrato, en el e-ésimo dominio.
- $S_{eh}^2$  = varianza muestral del h-ésimo estrato del e-ésimo dominio.
- $\bar{X}_{eh}$  = media muestral del h-ésimo estrato, del e-ésimo dominio.
- $Z_{\frac{\alpha}{2}}$  = factor de confiabilidad para un nivel de significancia  $\alpha$  dado.

## Anexo

### Marco y Muestra por Entidad Federativa

Entidad Federativa	Marco	Muestra	Mínima necesaria <sup>1</sup>
<b>Total</b>	<b>22 106</b>	<b>3 594</b>	<b>2 946</b>
1 Aguascalientes	381	93	75
2 Baja California	694	110	90
3 Baja California Sur	248	65	53
4 Campeche	364	60	50
5 Coahuila de Zaragoza	793	152	124
6 Colima	310	71	57
7 Chiapas	762	71	57
8 Chihuahua	673	147	120
9 Distrito Federal	2 061	358	300
10 Durango	395	70	57
11 Guanajuato	964	151	123
12 Guerrero	343	70	57
13 Hidalgo	349	108	87
14 Jalisco	1 525	216	180
15 México	1 069	111	91
16 Michoacán de Ocampo	707	95	77
17 Morelos	287	58	47
18 Nayarit	315	55	45
19 Nuevo León	1 535	246	204
20 Oaxaca	586	82	66
21 Puebla	759	124	100
22 Querétaro de Arteaga	487	82	67
23 Quintana Roo	366	73	60
24 San Luis Potosí	676	110	89
25 Sinaloa	772	97	80
26 Sonora	817	125	102
27 Tabasco	561	107	87
28 Tamaulipas	930	130	107
29 Tlaxcala	145	50	40
30 Veracruz de Ignacio de la Llave	1 165	130	110
31 Yucatán	800	106	86
32 Zacatecas	267	71	58

<sup>1</sup> Muestra mínima necesaria para la generación de resultados confiables.