

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

DIRECCIÓN GENERAL DE
ESTADÍSTICAS ECONÓMICAS

DIRECCIÓN DE MARCOS Y
MUESTREO

Encuesta Nacional de Empresas Constructoras Serie 2013

(ENEC Serie 2013)
- *Versión Preliminar* -

Documento Metodológico
sobre Diseño Muestral



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Índice

1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura sectorial	1
4. Unidad de Muestreo	1
5. Dominio de estudio	1
6. Marco de Muestreo	1
7. Estratificación	2
8. Esquema de muestreo	2
9. Tamaño de la muestra	2
10. Afijación de la muestra	3
11. Selección de la muestra	3
12. Cálculo de los factores de expansión	4
12.1. Ajuste por no respuesta	4
13. Estimadores	5

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

Está definida como 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 2013 (SCIAN 2013).

3. Cobertura Sectorial

La cobertura de la encuesta permite obtener estimaciones nacionales por subsector de actividad económica según el SCIAN 2013 y por entidad federativa.

4. Unidad de muestreo

La unidad de observación, es la empresa constructora.

5. Dominios de estudio

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

6. Marco de muestreo

El marco de muestreo se formó con base al Marco Estadístico Nacional de Unidades Económicas (MENUE) el cual se alimenta de los resultados definitivos de los Censos Económicos 2014 (CE 2014) actualizados por el Registro Estadístico de Negocios de México (RENEM) versión 2.0. El marco de muestreo está integrado por 25 297 empresas constructoras.

En el siguiente cuadro se presenta la participación porcentual por subsector de actividad económica del sector construcción.

Cuadro 1

Subsector	Descripción SCIAN 2013	Unidades Económicas		Personal Ocupado		Valor de Producción	
		Marco	PP(%)	Marco	PP(%)	(Miles de \$)	
						Marco	PP(%)
Total		25 297	100.00	764 928	100.00	325 844 466	100.00
236	Edificación	11 469	45.34	361 506	47.26	156 538 085	48.04
237	Construcción de obras de ingeniería civil u obra pesada	6 987	27.62	285 723	37.35	160 234 381	49.18
238	Trabajos especializados para la construcción	6 841	27.04	117 699	15.39	9 072 000	2.78

7. Estratificación

7.1 Formación del componente principal

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

7.2 Estratificación del componente principal

Se estableció una clasificación a cinco estratos, aplicando el algoritmo de k medias al componente principal, excepto para el primer estrato en el que se incluyeron con certeza a las empresas con más de 75 millones de pesos en el valor de la producción o personal ocupado mayor o igual a 250. Se obtuvo la siguiente distribución:

Cuadro 2

Estrato	Limites	Empresas
Total		25 297
1	Certeza	796
2	$2.09712 \leq CP < 10.02239$	1 654
3	$0.73162 \leq CP < 2.09712$	2 359
4	$- 0.20806 \leq CP < 0.73162$	3 841
5	$CP < - 0.20806$	16 647

8. Esquema de muestreo

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

a) Probabilístico

Las unidades de selección tienen una probabilidad conocida y distinta de cero de ser seleccionadas.

b) Estratificado

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

9. 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 12% 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.

σ_{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.

σ_{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 **3 782** empresas, considerando a las empresas incluidas con certeza (Ver anexo A).

10. Afijación de la muestra

Para los dominios de estudio, la muestra en cada estrato se asignó de acuerdo a la ponderación del componente principal por estrato 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.

σ_{eh} = desviación estándar del componente principal del h-ésimo estrato, de el e-ésimo dominio.

H = total de estratos por dominio.

La afijación de la muestra se presenta en el anexo A.

11. Selección de la muestra

Las empresas se seleccionaron de manera aleatoria, sin remplazo e independiente para cada dominio-estrato de acuerdo a la afijación. Se cuidó que el tamaño de muestra sea mínimo 4.

12. 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.

Los factores de expansión también se ajustaron con base en el siguiente concepto:

12.1 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 calculó 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.

13. Estimadores

El estimador del total de la característica X del e-ésimo dominio fue:

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

El estimador de la varianza fue:

$$\hat{V}(\hat{t}_e) = \sum_{h=1}^H N_{eh}^2 \left(1 - \frac{n_{eh}}{N_{eh}}\right) \frac{S_{eh}^2}{n_{eh}}$$

Donde:

$$S_{eh}^2 = \frac{1}{n_{eh}-1} \sum_{i=1}^{n_{eh}} [X_{ehi} - \bar{X}_{eh}]^2$$

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 α dado.

Anexo A

Marco y muestra de la Encuesta Nacional de Empresas Constructoras serie 2013 distribuida por entidad federativa.

Entidad		Marco	Muestra	Mínima Necesaria
Total		25 297	3 782	3 184
1	Aguascalientes	446	85	71
2	Baja California	822	118	100
3	Baja California Sur	355	76	62
4	Campeche	405	101	83
5	Coahuila de Zaragoza	845	140	118
6	Colima	369	76	62
7	Chiapas	1 053	112	92
8	Chihuahua	763	113	96
9	Ciudad de México	2 260	260	234
10	Durango	427	79	66
11	Guanajuato	1 210	141	120
12	Guerrero	342	95	78
13	Hidalgo	429	116	96
14	Jalisco	1 672	170	148
15	México	1 208	166	139
16	Michoacán de Ocampo	715	118	98
17	Morelos	293	86	70
18	Nayarit	340	101	83
19	Nuevo León	1 622	165	144
20	Oaxaca	825	138	111
21	Puebla	1 101	134	112
22	Querétaro de Arteaga	620	111	92
23	Quintana Roo	355	84	70
24	San Luis Potosí	738	126	105
25	Sinaloa	796	109	93
26	Sonora	1 008	121	104
27	Tabasco	601	97	81
28	Tamaulipas	974	133	112
29	Tlaxcala	205	65	52
30	Veracruz de Ignacio de la Llave	1 283	147	127
31	Yucatán	851	112	93
32	Zacatecas	364	87	72