

Instituto Nacional de Estadística y Geografía

Módulo de trayectorias laborales 2012

MOTRAL

Diseño muestral



INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Índice	Página
1. Objetivo de la encuesta	1
2. Población objetivo	1
3. Cobertura geográfica	1
4. Diseño de la muestra	1
4.1 Marco de la encuesta	1
5. Tamaño de la muestra	1
6. Selección de la muestra	1
7. Ajuste a los factores de expansión	2
7.1 Ajuste por No respuesta	2
7.2 Ajuste por proyección	2
8. Estimadores	3
9. Estimación de las precisiones	3
 Cuadro 1 Distribución de la muestra de personas por Grupo de Edad y Variable.	
 Cuadro 2 Distribución de la muestra de personas por Ciudad Autorrepresentadas y Variable.	

1. Objetivo de la encuesta

El objetivo del MOTRAL-2012 es generar información estadística sobre la trayectoria laboral, seguridad social y cultura de previsión social de la población.

2. Población objetivo

La encuesta está dirigida a las personas de 18 a 54 años de edad con experiencia laboral.

3. Cobertura geográfica

El módulo está diseñado para dar resultados para cada uno de los tres grupos de edad que corresponden a las 32 ciudades autorrepresentadas, estos grupos son:

- Personas de 18 a 29 años de edad.
- Personas de 30 a 44 años de edad.
- Personas de 45 a 54 años de edad.

En el cuadro 1, se presenta la distribución de la muestra de personas por Grupo de Edad y Variable.

4. Diseño de la muestra

El diseño del MOTRAL-2012 se caracteriza por ser probabilístico, lo que permite generalizar los resultados a toda la población de estudio. A la vez es estratificado, trietápico y por conglomerados, donde la unidad última de selección es la persona.

4.1 Marco de la encuesta

El marco de muestreo que se empleo para el MOTRAL-2012 fueron las viviendas con población de 18 a 54 años de edad, detectados en la Encuesta Nacional de Ocupación y Empleo (ENOE)¹ del mes de marzo de 2012.

5. Tamaño de la muestra

El tamaño de la muestra fue de 6 169 viviendas a nivel del agregado de las 32 áreas autorrepresentadas para obtener estimaciones de proporciones iguales o mayores del 10%, considerando una confianza del 90%, un error relativo máximo esperado del 15%, una tasa de No respuesta máxima esperada del 15%.

6. Selección de la muestra

En las viviendas que tuvieron más de una persona objeto de estudio se aplicó submuestreo, es decir se seleccionó sólo una persona con igual probabilidad, por lo tanto, la probabilidad de seleccionar de la k-ésima persona de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad se define de la siguiente manera:

$$P_{chijk} = P_{chij}^{ENOE} \frac{1}{m}$$

¹ Ver documento Metodológico del diseño muestral de ENOE.

Su factor de expansión es:

$$F_{chijk} = F_{chij}^{ENOE} m$$

Donde:

P_{chijk} = es la probabilidad de seleccionar de la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad, para el MOTRAL 2012.

P_{chij}^{ENOE} = es la probabilidad de seleccionar de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad, para la ENOE.

F_{chijk} = es el factor de expansión de la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad, para el MOTRAL 2012.

F_{chij}^{ENOE} = es el factor de expansión de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad.

m = número de personas de 18 a 54 años de edad con experiencia laboral por vivienda.

En el cuadro 2, se presenta la Distribución de la muestra de personas por Ciudad Autorepresentadas y Variable.

7. Ajuste a los factores de expansión

Los factores de expansión elaborados conforme al procedimiento antes descrito se ajustan en base a los siguientes conceptos:

7.1 Ajuste por No respuesta

El ajuste por No respuesta, se realizó para las personas seleccionadas a nivel UPM, en cada uno de los estratos, mediante la siguiente expresión:

$$F'_{chijk} = F_{chijk} \frac{m_{hi}}{m_{hi}^*}$$

Donde:

F'_{chijk} = factor de expansión corregido por No respuesta, de la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad.

m_{hi} = número de personas seleccionadas de 18 a 54 años de edad con experiencia laboral, en la i-ésima UPM, en el h-ésimo estrato.

m_{hi}^* = número de personas seleccionadas de 18 a 54 años de edad con experiencia laboral, de las cuales se obtuvo respuesta, en la i-ésima UPM, en el h-ésimo estrato.

F_{chijk} = factor de expansión de la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad.

7.2 Ajuste por proyección

Los factores de expansión ajustados por la No respuesta se corrigen, a fin de asegurar que en cada dominio de interés de la encuesta se obtenga la población total determinada por la proyección de población generada por INEGI referida al punto medio del levantamiento, mediante la siguiente expresión:

$$F_D'' = F_D^* \frac{PROy_D}{PEXP_D}$$

Donde:

- F_D'' = factor de expansión corregido por proyección en el dominio D.
 F_D^* = factor de expansión corregido por No respuesta en el dominio D.
 $PROy_D$ = población en el dominio D, según proyección.
 $PEXP_D$ = población total a la que expande la encuesta en el dominio D.

8. Estimadores

El estimador del total de la característica X es:

$$\hat{X} = \sum_c \sum_h \sum_i F_{chijk} \left(\sum_j \sum_k X_{chijk} \right)$$

Donde:

- F_{chijk} = factor de expansión final de la k-ésima persona, de la j-ésima vivienda, de la i-ésima UPM, del h-ésimo estrato, de la c-ésima ciudad.
 X_{chijk} = valor observado de la característica de interés X en la k-ésima persona, en la j-ésima vivienda, en la i-ésima UPM, en la h-ésimo estrato, en la c-ésima ciudad.

Para la estimación de proporciones, tasas y promedios se utiliza el estimador de razón:

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}$$

Donde, \hat{Y} se define en forma análoga a \hat{X} .

9. Estimación de las precisiones

Para la evaluación de los errores de muestreo de las principales estimaciones nacionales se usó el método de Conglomerados Últimos², basado en que la mayor contribución a la varianza de un estimador, en un diseño trietápico es la que se presenta entre las unidades primarias de muestreo (UPM), el término "Conglomerados Últimos" se utiliza para denotar el total de unidades en muestra de una unidad primaria de muestreo.

Para obtener las precisiones de los estimadores de razón, conjuntamente al método de Conglomerados Últimos se aplicó el método de series de Taylor, obteniéndose la siguiente fórmula para estimar la precisión de \hat{R} :

$$\hat{V}(\hat{R}) = \frac{1}{\hat{Y}^2} \sum_h \left\{ \sum_c \frac{k_{ch}}{k_{ch}-1} \sum_i \left[\left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right) - \hat{R} \left(\hat{Y}_{chi} - \frac{1}{k_{ch}} \hat{Y}_{ch} \right) \right]^2 \right\}$$

² Vease Hasen, M. H. Horwitz, W.N. y Madow, W.G., Sample Survey Methods and Theory, (1953) Vol. 1 pág. 242.

Donde:

\hat{X}_{chi} = total ponderado de la variable de estudio X en la i-ésima UPM, en el h-ésimo estrato, en la c-ésima ciudad.

\hat{X}_{ch} = total ponderado de la variable de estudio X en el h-ésimo estrato, en la c-ésima ciudad.

k_{ch} = número de UPM en el h-ésimo estrato, en la c-ésima ciudad.

Estas definiciones son análogas para la variable de estudio Y.

La estimación de la varianza del estimador de un total, se calcula con la siguiente expresión:

$$\hat{V}(\hat{X}_{NAL}) = \sum_{c=1}^{32} \sum_{h=1}^{L_c} \frac{k_{ch}}{k_{ch}-1} \sum_{i=1}^{k_{ch}} \left(\hat{X}_{chi} - \frac{1}{k_{ch}} \hat{X}_{ch} \right)^2$$

Las estimaciones de la desviación estándar (D.E.), efecto de diseño (DEFF) y coeficiente de variación (C.V.) se calculan mediante las siguientes expresiones:

$$D.E. = \sqrt{\hat{V}(\hat{\theta})}$$

$$DEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}(\hat{\theta})_{MAS}}$$

$$C.V. = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

Donde:

$\hat{\theta}$ = estimador del parámetro poblacional θ .

$\hat{V}(\hat{\theta})_{MAS}$ = estimador de la varianza bajo muestreo aleatorio simple.

Finalmente, el intervalo de confianza al $100(1-\alpha)\%$, se calcula de la siguiente forma:

$$I_{1-\epsilon} = \left(\hat{\theta} - z_{\alpha/2} \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{\alpha/2} \sqrt{\hat{V}(\hat{\theta})} \right)$$

Distribución de la muestra de personas por Grupo de Edad y Variable

Cuadro 1

Grupo de Edad	Población Ocupada	Población en Desempleo Abierto	Población No Económicamente Activa	Total
18-29 años	1 387	624	370	2 381
30-44 años	1 774	336	376	2 486
45-54 años	923	137	242	1 302
Total	4 084	1 097	988	6 169

Distribución de la muestra de personas por Ciudad Autorrepresentadas y Variable

Cuadro 2

Área Metropolitana	Población Ocupada	Población en Desempleo Abierto	Población No Económicamente Activa	Total
01 México	106	44	24	174
02 Guadalajara	201	44	35	280
03 Monterrey	212	55	38	305
04 Puebla	113	45	38	196
05 León	192	58	46	296
07 San Luis Potosí	135	27	25	187
08 Mérida	125	22	28	175
09 Chihuahua	142	28	25	195
10 Tampico	114	27	34	175
12 Veracruz	119	19	32	170
13 Acapulco	73	29	29	131
14 Aguascalientes	158	41	33	232
15 Morelia	92	24	28	144
16 Toluca	170	66	38	274
17 Saltillo	191	42	32	265
18 Villahermosa	88	35	28	151
19 Tuxtla Gutiérrez	91	36	30	157
21 Tijuana	174	39	36	249
24 Culiacán	139	33	34	206
25 Hermosillo	158	33	33	224
26 Durango	110	41	29	180
27 Tepic	99	30	28	157
28 Campeche	112	21	22	155
29 Cuernavaca	90	20	27	137
31 Oaxaca	87	28	32	147
32 Zacatecas	124	41	39	204
33 Colima	122	29	29	180
36 Querétaro	166	18	31	215
39 Tlaxcala	60	47	30	137
40 La Paz	110	26	22	158
41 Cancún	140	25	25	190
43 Pachuca	71	24	28	123
Nacional	4 084	1 097	988	6 169