



El proyecto *Diagnóstico y Políticas Públicas Estatales para retrasar el Primer Consumo de Drogas Ilegales entre las Juventudes* tiene como objetivo identificar perfiles de consumo de drogas ilegales entre jóvenes de 12 a 17 años de edad en 10 entidades federativas con mayor consumo de drogas ilegales.

Los objetivos específicos del estudio se definieron como sigue:

- a. Identificar factores de riesgo y protección relacionados con el consumo inicial temprano de drogas ilegales
- b. Proponer líneas de trabajo para realizar políticas públicas estatales que lo retrasen y prevengan.
- c. Realizar un diagnóstico, a través de una encuesta, en 10 de las entidades federativas con mayor uso de drogas ilegales en las y los jóvenes de entre 12 y 17 años, a saber: Aguascalientes, Ciudad de México, Coahuila de Zaragoza, Durango, Guanajuato, Hidalgo, Morelos, Oaxaca, Quintana Roo y Yucatán.
- d. Identificar los perfiles sociodemográficos y epidemiológicos de esa población, para fortalecer los correspondientes factores de protección.
- e. Generar recomendaciones para políticas públicas estatales que sirvan para postergar el consumo inicial temprano de drogas ilegales entre jóvenes.

Para cumplir estos objetivos, uno de los elementos de la propuesta original fue el levantamiento de información en 10 entidades federativas a razón de 600 casos en cada una de ellas, lo que permitiría un margen de error de +/-4% por entidad.

Una de las dudas más frecuentes que se tiene al momento de analizar los datos es determinar si el número de observaciones es suficiente para reflejar el comportamiento de la población real. En la literatura del muestreo no existe un número mínimo de casos a partir del cual se pueda inferir que la muestra es significativa (Escobar, 2013; Johnson & Elliott, 1998; Lumley, 2010; Martínez, 2017). A pesar de ello, la convención suele ser que los datos mínimos para determinar que una muestra es representativa debe mantener un margen de error menor o igual a +/-5% y un nivel de confianza del 90%, de hecho, las pruebas estadísticas que lleva a cabo el Instituto Nacional de Estadística y Geografía se hacen con los parámetros antes mencionados.

En este sentido, para garantizar la representatividad de la muestra, se calcula el **margen de error**, que se obtiene a partir de despejar la siguiente expresión:

$$n = \frac{N\sigma^2 Z_{\alpha}^2}{(N-1)e^2 + \sigma^2 Z_{\alpha}^2}$$

Al despejar e (margen de error) la ecuación resultante es:

$$e = \sqrt{\frac{\sigma^2 Z_{\alpha}^2}{n}}$$

Así, se considera un nivel de confianza $\alpha = 95\%$ y asumiendo que Z_{α} representa el cuartil de la distribución normal que acumula una distribución de α su valor en tablas es de **1.96**. Además, para estimar σ^2 se asume que la proporción es el **50%** y n es el número total de encuestas realizadas. Con estos parámetros, los resultados son los siguientes:

Entidad	INEGI. Censo de Población y Vivienda 2010	Encuestas realizadas	Encuestas esperadas	Margen de error (%)
Aguascalientes	147,108	573	600	4.09
Ciudad de México	821,159	431	600	4.72
Coahuila de Zaragoza	318,870	559	600	4.14
Durango	201,248	333	600	5.37
Guanajuato	694,768	568	600	4.11
Hidalgo	324,409	93	600	
Morelos	207,305	28	600	
Oaxaca	507,486	602	600	3.99
Quintana Roo	147,719	551	600	4.17
Yucatán	225,775	420	600	4.78

Cabe mencionar que para los estados de Hidalgo y Morelos no se calcula el error debido a que el levantamiento aún está incompleto y se va a retomar el operativo de campo para terminar con la recolección de la muestra. Por otro lado, para el estado de Durango aún se valorará la pertinencia de levantar más encuestas por lo que su margen de error podría disminuir, a pesar de que el valor actual está dentro de los márgenes aceptados. Además mencionar que a nivel general, considerando las 4,158 el margen de error es de **1.52%**.

Siguientes pasos

Si bien es cierto que el cálculo anterior confirma que la muestra es confiable, ahora se propone evaluar los estimadores, es decir, datos específicos. De acuerdo Cochran (2007) para saber si un dato es estadísticamente significativo es necesario evaluarlo con base en los criterios que diseño la muestra. De hecho, Molina (2019) advierte que la significancia no está ligada con un número en particular, sino en la forma en que se levantan los datos, ya que se pueden levantar muchas encuestas pero si solo se ubican en una misma ubicación, difícilmente los resultados serán confiables. Esto significa que el número de casos dentro de una encuesta por sí mismo no garantiza que la información sea de buena calidad.

Ante esta situación, lo más relevante es garantizar la calidad de los datos. Para ello no existe un método unánime que sirva de manera universal; sin embargo, una de las recomendaciones que hace Naciones Unidas y que sigue el INEGI es calcular el coeficiente de variación (INEGI, 2011; ONU, 2010). Se trata de un estadístico que contrasta el tamaño de la media y la variabilidad de la variable, por lo que ofrece el grado de variabilidad de los datos:

$$C_V = \frac{\sigma}{\bar{x}}$$

Los parámetros que utiliza¹ INEGI (2011) para determinar si un dato es confiable -o no- son los siguientes: si el coeficiente de variación es menor a 15%, el dato es de buena calidad; si está entre 15% y 25%, entonces su calidad es aceptable; si supera 25%, entonces tienen baja calidad. Bajo esta lógica, se llevaron a cabo este ejercicio y la estimación general del margen de error para todos los estados y los resultados son los siguientes:

Cuadro 1. Coeficiente de variación para la población total, hombres y mujeres

Entidad	TOTAL						
	Censo de Población y Vivienda 2010. Total	Estimación (datos SIMO)	Variación (%)	Error estándar	Coeficiente de variación	Límite (95%)	
						Inferior	Superior
Aguascalientes	147,108	144,509	1.8%	16421	11.40	112313	176705
Ciudad de México	821,159	803,187	2.2%	36556	4.55	731511	874863
Coahuila de Zaragoza	318,870	311,631	2.3%	23444	7.52	265663	357599
Guanajuato	694,768	691,937	0.4%	116703	16.90	463112	920762
Oaxaca	507,486	501,406	1.2%	48458	9.66	406227	596585
Quintana Roo	147,719	148,477	0.5%	22758	15.30	103855	193099
Yucatán	225,775	219,931	2.6%	28834	13.10	163394	276468

¹ Para ver los criterios de INEGI ir a:

<https://www.inegi.org.mx/programas/enoe/15ymas/default.html#Documentacion>, dar click en Documentación/Material de apoyo y descargar el documento "Pruebas de significancia estadística. Pruebas de hipótesis"

HOMBRES

Entidad	Censo de Población y Vivienda 2010.	Estimación	Variación (%)	Error estándar	Coeficiente de variación	Límite (95%)	
						Inferior	Superior
Aguascalientes	73,918	72,954	1.3%	10775	14.80	51827	94081
Ciudad de México	414,328	404,647	2.3%	28734	7.10	348308	460986
Coahuila de Zaragoza	161,265	157,578	2.3%	11850	7.52	134344	180812
Guanajuato	348,702	346,871	0.5%	31983	9.22	284161	409581
Oaxaca	254,289	251,743	1.0%	28476	11.30	195807	307679
Quintana Roo	74,849	76,189	1.8%	16197	21.30	44431	107947
Yucatán	114,386	113,264	1.0%	14843	13.10	84161	142367

MUJERES

Entidad	Censo de Población y Vivienda 2010.	Estimación	Variación (%)	Error estándar	Coeficiente de variación	Límite (95%)	
						Inferior	Superior
Aguascalientes	73,190	71,555	2.2%	13214	18.50	45646	97464
Ciudad de México	406,831	398,540	2.0%	32439	8.14	334935	462145
Coahuila de Zaragoza	157,605	154,053	2.3%	22790	14.80	109367	198739
Guanajuato	346,066	345,066	0.3%	113434	22.90	122652	567480
Oaxaca	253,197	249,663	1.4%	52261	20.90	147015	352311
Quintana Roo	72,870	72,288	0.8%	17838	24.70	37313	107263
Yucatán	111,389	106,667	4.2%	32451	24.40	43040	170294

* La prueba se hizo con un nivel de confianza del 95%, INEGI hace las pruebas al 90%

Esto corrobora que los datos están dentro de los parámetros de buena calidad y aceptable, con lo cual se infiere que la información si refleja el comportamiento de la población real.

Otro aspecto que se debe hacer notar es para llegar a este resultado se llevó a cabo un proceso de calibración de los factores de expansión. Esto es un paso crucial, tal y como se explican Valliant y Dever (2018) ya que sirve para dar representatividad a la encuesta que se está utilizando. En ese sentido, en la columna de Variación (%) se puede observar que todas las variaciones, es decir, la diferencia entre los datos de la encuesta de SIMO y los datos de INEGI son menores a 5%.

Referencias

- Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.
- Escobar, E. (2013). *On variance estimation under complex sampling designs*.
- INEGI. (2011). *Diseño de la muestra en proyectos de encuesta*. Instituto Nacional de Estadísticas y Geografía Aguascalientes-México.
- Johnson, D. R., & Elliott, L. A. (1998). Sampling Design Effects: Do They Affect the Analyses of Data from the National Survey of Families and Households? *Journal of Marriage and Family*, 60(4), 993–1001. <https://doi.org/10.2307/353640>
- Lumley, T. (2010). *Complex surveys : a guide to analysis using R*. John Wiley. <http://pbidi.unam.mx:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat02025a&AN=lib.MX001001689915&lang=es&site=eds-live>
- Martínez Sánchez, J. C. (2017). Una aproximación metodológica al uso de datos de encuestas en hogares. *Realidad, Datos y Espacio. Revista Internacional de Estadística y Geografía*, 8(2), 58–80.
- Molina, I. (2019). Desagregación de datos en encuestas de hogares. *CEPAL - Serie Estudios Estadísticos*, 97.
- ONU. (2010). *Principios y recomendaciones para los censos de población y habitación. Revisión 2*. Naciones Unidas Nueva York.
- Valliant, R., & Dever, J. A. (2018). *Survey weights: a step-by-step guide to calculation*. Stata Press College Station, TX.