



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS DE COMPONENTES PRINCIPALES
DE LA ENCUESTA NACIONAL DE VICTIMIZACIÓN
Y PERCEPCIÓN SOBRE SEGURIDAD PÚBLICA

TESIS

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A

CECILIA BERENICE LÓPEZ PÉREZ

Tutora

ACT. ADRIANA VELÁZQUEZ RAMÍREZ

Ciudad Universitaria, noviembre 2022





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos de la alumna

López

Pérez

Cecilia Berenice

301124058

Universidad Nacional Autónoma de México

Facultad de Ciencias

5527614987

2. Datos de la Tutora

Actuaria

Adriana

Velázquez

Ramírez

3.- Sinodal 1

Maestro en Ciencias

Rodrigo

Quijón

Hipólito

4. Sinodal 2

Maestro en Administración

Luis Alberto

Galván

Contreras

5. Sinodal 3

Maestra en Ciencias

Jessika Dilhery

Lucas

Flores

6. Sinodal 4

Licenciado en Economía

Edgar Arturo

Durán

Loera

Título

Análisis de componentes principales de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública

73 Páginas

2022

Índice general

Figuras.....	I
Cuadros	III
Introducción	1
Capítulo 1. La victimización y percepción sobre seguridad pública.....	3
1.1 Antecedentes	3
1.2 Descripción del fenómeno.....	4
1.2.1 Victimización.....	5
1.2.2 Seguridad Pública	5
1.3 Contexto de la ENVIPE 2020.....	6
1.3.1 Implementación de políticas públicas.....	9
1.3.2 Prácticas internacionales para la medición.....	10
1.3.3 Captación de información durante la contingencia sanitaria (COVID-19)	13
1.4 Diseño estadístico.....	13
1.4.1 Objetivos específicos	13
1.4.2 Diseño muestral.....	14
1.4.3 Población objetivo	15
1.4.4 Levantamiento y cobertura	15
1.5 Análisis descriptivo	15
1.5.1 Prevalencia delictiva en hogares.....	16
1.5.2 Prevalencia delictiva en las personas.....	17
1.5.3 Incidencia delictiva	17
1.5.4 Cifra Negra.....	26
1.5.5 Percepción sobre seguridad pública	28
1.6 Conclusión del capítulo.....	31
Capítulo 2. Modelo factorial para el análisis multivariado.....	32
2.1 Introducción	32
2.2 Planteamiento del análisis factorial.....	32
2.2.1 Componentes principales	32
2.2.2 Diseño de la investigación	33
2.3 Objetivos específicos	35
2.3.1 Reducir información	35
2.4 Etapas del análisis factorial.....	35
2.4.1 Análisis factorial exploratorio	36
2.4.1.1 Estandarización de variables.....	37
2.4.1.2 Matriz de correlación reproducida	38
2.4.2 Análisis factorial confirmatorio	39
2.4.2.1 Hipótesis y supuestos.	39
2.4.2.2 Test de esfericidad de Bartlett.....	40
2.4.2.3 Medida de la adecuación muestral.....	41
2.4.2.4 Regla de KMO	41
2.4.2.5 Criterio de porcentaje de varianza.....	42
2.4.2.6 Extracción de factores	43

2.4.2.7	Determinación del número de factores.....	43
2.4.2.8	Cálculo Puntuaciones factoriales.....	44
2.5	Validación del modelo.....	45
2.6	Conclusión del capítulo.....	46
Capítulo 3. Aplicación del modelo factorial a la ENVIPE 2020.....		47
3.1	Introducción.....	47
3.2	Diseño de la investigación.....	47
3.2.1	Reducir información.....	48
3.2.2	Objetivo específico.....	49
3.2.3	Integración de fuentes de datos.....	49
3.2.4	Herramienta de explotación.....	49
3.3	Elección del método estadístico.....	51
3.3.1	Diseño metodológico.....	51
3.3.2	Hipótesis y variables.....	52
3.3.3	Criterios y estandarización de variables.....	55
3.3.4	Matriz de correlación.....	55
3.3.5	Factores en la solución.....	58
3.4	Construcción del índice de victimización.....	62
3.5	Conclusión del capítulo.....	67
Conclusiones.....		68
Bibliografía.....		69

Figuras

Figura 1.1 Cuadro comparativo de encuestas de victimización a nivel internacional	12
Figura 1.2 Prevalencia delictiva en hogares.....	16
Figura 1.3 Prevalencia delictiva en las personas	17
Figura 1.4 Incidencia delictiva en las personas	17
Figura 1.5 Incidencia delictiva en las personas por sexo	18
Figura 1.6 Incidencia por tipo de delito.....	19
Figura 1.7 Incidencia delictiva por sexo.....	19
Figura 1.8 Incidencia delictiva por entidad federativa.....	20
Figura 1.9 Incidencia delictiva en las personas	20
Figura 1.10 Incidencia delictiva por robo o asalto en la calle o en el transporte público	21
Figura 1.11 Incidencia delictiva por robo o asalto en la calle o transporte según sitio de ocurrencia ...	22
Figura 1.12 Incidencia delictiva por robo o asalto en calle o transporte público en la CDMX	22
Figura 1.13 Incidencia delictiva por extorsión	23
Figura 1.14 Incidencia delictiva por secuestro de algún integrante del hogar.....	23
Figura 1.15 Incidencia delictiva en las personas	24
Figura 1.16 Incidencia de delitos más frecuentes	25
Figura 1.17 Porcentaje de delitos donde la víctima estuvo presente.....	25
Figura 1.18 Porcentaje de delitos cometidos con portación de arma	26
Figura 1.19 Porcentaje de delitos denunciados en el Ministerio Público	26
Figura 1.20 Cifra Negra.....	27
Figura 1.21 Resultados de la carpeta de investigación.....	27
Figura 1.22 Tazones para no denunciar delitos	28
Figura 1.23 Distribución porcentual sobre los temas que generan mayor preocupación	28
Figura 1.24 Percepción de la inseguridad y de la seguridad pública	29
Figura 1.25 Percepción de la inseguridad por entidad federativa.....	29
Figura 1.26 Distribución porcentual de la población que manifiesta sentirse insegura en espacios públicos o privados	30
Figura 1.27 Distribución porcentual de la población que manifiesta sentirse insegura en espacios públicos o privados por sexo.....	30
Figura 2.1 Pasos necesarios para realizar un análisis factorial.....	34
Figura 3.1 Pasos necesarios para realizar un análisis factorial.....	48

Figura 3.2 Fuentes de datos para exportar en SPSS® 50

Cuadros

Cuadro 1.1 Tasa de víctimas de delito por cada 100 mil habitantes por año	10
Cuadro 1.2 Porcentaje de población que percibe inseguridad en el Municipio o Alcaldía.....	10
Cuadro 1.3 Diseño estadístico	13
Cuadro 1.4 Incidencia delictiva 2018-2019.....	18
Cuadro 1.5 Incidencia delictiva según dominio 2018-2019	21
Cuadro 1.6 Incidencia delictiva por secuestro de algún integrante del hogar.....	24
Cuadro 3.1 Descripción de variables.....	53
Cuadro 3.2 Matriz de Correlación.....	57
Cuadro 3.3 Prueba de KMO y Barlett.....	59
Cuadro 3.4 Matriz de Correlación anti-imagen.....	60
Cuadro 3.5 Prueba de esfericidad de Bartlett	61
Cuadro 3.6 Matriz de componentes	62
Cuadro 3.7 Comunalidades.....	63
Cuadro 3.8 Varianza total explicada	64
Cuadro 3.9 Matriz de componentes rotado	64
Cuadro 3.10 Matriz de coeficiente de puntuación de componente rotado	65
Cuadro 3.11 Resultados de medias, varianzas y desviaciones estándar de los 15 indicadores	66

Introducción

En este estudio se abordará el análisis de la victimización y seguridad pública en el contexto de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE 2020) para el diseño e implementación de políticas públicas, las prácticas internacionales para medir la victimización y el entorno de la contingencia sanitaria por el coronavirus (COVID-19). Este análisis abordará esta problemática con la medición de la victimización del hogar y personal, número de víctimas, delitos, cifra negra y la percepción sobre la seguridad en el lugar donde viven o realizan sus actividades cotidianas.

En esta tesis se aplicará un análisis de componentes principales, que es un análisis multivariado y sus orígenes se asocian con el análisis factorial de Francis Galton y su concepto de factores latentes, posteriormente Karl Pearson extiende las ideas de Galton. Afifi, May & Clark en su libro *Practical Multivariate Analysis* establecen que el método permite simplificar relaciones complejas y diversas que existen entre un conjunto de variables observadas y que se puede resumir como un método de transformación de variables originales dentro de una nueva variable. La nueva variable es llamada componente principal. Cada componente principal es la combinación lineal de la variable original.

En términos generales, los componentes principales analizan la estructura de las correlaciones entre un gran número de variables mediante la definición de un conjunto de dimensiones subyacentes comunes conocidos como factores. A partir de los resultados de la ENVIPE 2020 se identificará primero las dimensiones separadas por la estructura y después se determinará el grado en que se explica cada variable por cada dimensión.

La formulación del análisis factorial es muy similar al análisis de componentes principales, ambas son técnicas estadísticas del análisis multivariado, para identificar el menor número de factores que pueden ser usados para representar de forma resumida la información contenida en un conjunto numeroso de variables interrelacionadas, pero conceptualmente tienen diferentes significados. Las diferencias entre el análisis factorial (AF) y los componentes principales (CP) es si la matriz de correlación existe, las CP también existen, mientras que el modelo AF podría ser aceptado o no mediante un test estadístico.

La tesis se presenta grosso modo en tres capítulos:

El objetivo del capítulo 1 es mostrar un panorama sobre la violencia a partir del contexto sobre el cual se diseñó y levantó la ENVIPE en el año 2020; así como el análisis descriptivo de sus resultados sobre la prevalencia e incidencia delictiva en hogares y personas, cifra negra y la percepción sobre seguridad pública.

Por su parte, el objetivo del capítulo 2 es describir el marco metodológico del análisis de componentes principales como técnica de análisis multivariante para obtener los componentes o nuevas variables que cumplan con el objetivo del análisis que en particular es la medición de la victimización y percepción sobre seguridad pública. Para este fin, se planteará el problema y se mostrarán los aspectos teóricos para la extracción de los componentes principales, su interpretación gráfica y la inferencia estadística avanzada.

Finalmente, el objetivo del capítulo 3 es aplicar el método de componentes principales para determinar el nivel de victimización y percepción sobre seguridad pública, los cuales permitirán determinar los componentes y construcción del índice de victimización.

Capítulo 1.

La victimización y percepción sobre seguridad pública

1.1 Antecedentes

La realización de encuestas de victimización han cobrado en los últimos tiempos a nivel internacional y nacional una importancia relevante, tanto en el ámbito de los gobiernos y las autoridades responsables de las funciones vinculadas con la seguridad y la justicia como en el ámbito académico; a ello ha contribuido, sin duda, la importancia que actualmente conceden las sociedades a la aspiración de un ambiente de vida sin violencia, que no atente contra la integridad física y patrimonial de los ciudadanos y de las instituciones.

En el año 2011 se crea la *Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública* (ENVIPE) que forma parte de los proyectos estadísticos del Subsistema Nacional de Información de Gobierno, Seguridad Pública e Impartición de Justicia (SNIGSPIJ), pues se reconoce la necesidad dentro del Estado Mexicano de producir, integrar, administrar y difundir información estadística y geográfica de interés nacional de calidad, pertinente, veraz y oportuna relacionada con la gestión y desempeño de las instituciones públicas de los tres ámbitos de gobierno y de los tres Poderes de la Unión, específicamente sobre los temas de gobierno, seguridad pública e impartición de justicia.

Se decidió que la ENVIPE se realizara de forma anual y que quedaría a cargo de Instituto Nacional de Estadística y Geografía (INEGI). Es entonces que del 14 de marzo al 22 de abril de 2011 se levanta la primera encuesta, misma que atiende las recomendaciones de la Oficina de Naciones Unidas contra la Droga y el Delito (UNODC), de los principales usuarios de información, de las autoridades en materia de seguridad pública, de expertos académicos y de la experiencia previa del INEGI. Primeramente, se identificó la necesidad de actualizar la información sobre la percepción que tienen los habitantes donde viven y la evaluación de las autoridades policiacas; así como de obtener datos que permitan estimar los hechos delictivos denunciados y no denunciados en todo el país.

La información es obtenida a nivel nacional, estatal y de 17 zonas urbanas que permiten realizar estimaciones de la incidencia delictiva que afecta a los hogares y a las personas miembros de los hogares mayores de 18 años de edad, durante el 2010; así como las características del delito, las víctimas y el contexto de la victimización, buscando obtener información sobre la percepción de la seguridad pública, la percepción del desempeño y experiencias con las instituciones a cargo de la seguridad pública y la justicia. Esto con el fin de proveer información al público en general y generar elementos para la toma de decisiones de política pública en estas materias.

En virtud de su importancia, desde diciembre de 2011 la ENVIPE fue declarada información de interés nacional por la Junta de Gobierno del INEGI, lo cual significa que su uso es obligatorio para la Federación, los Estados, la Ciudad de México y los municipios.

La serie estadística ENVIPE responde a este entorno, recabando información; teniendo como objetivos específicos:

- Hacer estimaciones sobre el número de personas que sufrieron victimización delictiva.
- Calcular el número de delitos ocurridos.
- Hacer estimaciones sobre la “cifra negra” de los delitos y sus causas.

- Medir la percepción actual de la población de 18 años y más sobre la seguridad del lugar donde viven (colonia o localidad, municipio, entidad federativa) y sobre los lugares en donde realizan sus actividades cotidianas (trabajo, escuela, mercado, transporte público, entre otros).
- Medir el grado de confianza en las instituciones de seguridad pública y la percepción sobre el desempeño de las autoridades de seguridad pública y justicia.
- Identificar los cambios en las actividades de las personas por temor al delito.
- Estimar los gastos de los hogares para protegerse de la delincuencia.
- Conocer la repercusión del delito sobre las víctimas.
- Reunir información sobre las características sociodemográficas de las víctimas.
- Reunir información sobre las características del victimario y su relación con la víctima.
- Identificar actitudes y experiencias de las víctimas con las instituciones de seguridad pública y de procuración de justicia.

La ENVIPE 2020 es el décimo ejercicio de esta serie, a la cual anteceden los ejercicios anuales realizados de manera ininterrumpida.

1.2 Descripción del fenómeno

El bienestar de las personas y las comunidades se ve perjudicado por la actividad criminal en sus distintas formas. Aquellos que se ven directamente perjudicados por el delito sufren afectaciones económicas, sociales o incluso psicológicas, mientras que el temor al delito incide y restringe la vida de las personas de varias maneras.

La situación delictiva que vive un país influye en la calidad de vida de las personas; por lo tanto, la generación de información basada en evidencia contribuye al proceso de políticas públicas en la materia. Entre 1936 y 1940 México era el país del mundo con la mayor tasa de homicidios, culminando en 1940 con una tasa de 67.4 asesinatos por cada 100 000 habitantes. Sin embargo, a partir de este lapso la tasa empezó a declinar y, así, en 1950 fue de 48 por 100 000; en 1960 de 31; en 1970 de 17; en 1980 y en 1990 de 18 y para el 2000 de 11 por cada 100 000, hasta llegar en 2005 a 9.49 (Menéndez, 2012).

Esta tendencia ha sido constante, pese a los notorios incrementos ocurridos en la primera mitad de los noventa y entre 2008 y 2010, pero sin llegar a los niveles que observamos entre 1936 y 1960, ya que, por ejemplo, en el 2009 la tasa fue de 17.88 asesinados por cada 100 000 habitantes. Fue en la década de 1960 se identificaron las debilidades de las fuentes provenientes de registros administrativos, por lo que se buscaron métodos alternativos que ofrecieran una caracterización integral de la situación delictiva en un país.

Las encuestas de victimización surgen como proyectos estadísticos que complementan a los registros administrativos, generando estimaciones sobre los delitos que afectan a las personas y que no son denunciados ante alguna autoridad (cifras negras), contexto de la victimización, percepción de inseguridad, desempeño institucional, entre otros. Los registros administrativos son una fuente de información muy valiosa para conocer los índices delictivos de un país, sin embargo, no ofrecen un análisis exhaustivo sobre el fenómeno de la criminalidad, ya que únicamente cuentan con información sobre delitos que son denunciados ante la autoridad correspondiente.

Ante la situación de inseguridad en México se concreta en 2010 un acuerdo de colaboración con la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC, por sus siglas en inglés), para instituir el Centro de Excelencia (CdE) para Información Estadística de Gobierno, Seguridad Pública, Victimización y Justicia. El propósito de este Centro es generar nuevas metodologías que permitan mejorar la calidad de la medición de la delincuencia, así como coordinar esfuerzos nacionales o en conjunto con la comunidad internacional para obtener estadísticas e indicadores comparables.

El CdE apoya en el desarrollo de estándares internacionales para generar información estadística de gobierno, seguridad pública, victimización y justicia; además contribuye con la ENVIPE mediante la generación, desarrollo y homologación de información estadística de calidad sobre el delito, la violencia, la seguridad pública en países de América Latina y el Caribe.

1.2.1 Victimización

De acuerdo con el INEGI, *victimización* se define como un delito que afecta a una persona o a un hogar (ENVIPE 2020). Este es el tema central de la encuesta y se enfoca en las experiencias de los hogares y los individuos frente al delito. Se determinan las experiencias de victimización, se indaga acerca de las características de los delitos y la situación de la denuncia, de esta forma se puede obtener una imagen completa del hecho delictivo.

Para medir la victimización se utilizan los siguientes criterios:

- Incidentes delictivos en el hogar: son los delitos que pudieron haber ocurrido y que afectaron a todos los integrantes del hogar; en este sentido se considera al hogar como la víctima.
- Experiencia de victimización personal: son los delitos que afectaron directamente a quien lo está reportando y puede proporcionar detalles acerca del delito.
- Contexto de victimización: a los informantes que reportan haber sido víctimas directas de algún delito, se les pregunta sobre los detalles del incidente como características del (los) delincuente(s), uso de armas y/o violencia física.
- Denuncia ante la autoridad correspondiente: se determina si el delito del cual fue víctima el informante fue denunciado o no ante el Ministerio Público y cuál es el estatus de la denuncia, o si el delito no fue denunciado y las razones de la no denuncia.
- Experiencia al denunciar: se indaga la manera en que el Ministerio Público se ocupó de la denuncia y el tipo de ayuda recibida.
- Daños y costos de la victimización: se analizan los daños y pérdidas a consecuencia del delito, pueden ser físicas, emocionales, psicológicas y económicas.
- Otros detalles de la victimización: se explora sobre los detalles más relevantes por cada tipo de delito; por ejemplo, para robo en casa habitación, asalto en la calle o transporte público, se explora por el tipo de objetos robados. Para el delito de fraude se profundiza por el tipo de fraude que sufrió la víctima.

1.2.2 Seguridad Pública

De acuerdo al Artículo 21 de la Constitución Política de los Estados Unidos Mexicanos, la Seguridad Pública es la función a cargo de la Federación, las entidades federativas y los Municipios, que comprende la prevención de los delitos; la investigación y persecución para hacerla efectiva.

La Seguridad Pública forma parte esencial del bienestar de una sociedad. La Federación debe generar las condiciones que permitan al individuo realizar sus actividades cotidianas con la confianza de que su vida, su patrimonio y sus otros bienes están exentos de todo peligro, daño o riesgo.

Por ello, la ENVIPE tiene por objetivo medir la percepción sobre seguridad pública y su relación con el delito; esto es de gran importancia para conocer la realidad que vive la población ya que una percepción desfavorable sobre la seguridad pública incide negativamente en ella; además que la confianza disminuye en las instituciones encargadas de la seguridad pública. Para cumplir este objetivo es necesario abordar la temática desde distintas perspectivas, las categorías conceptuales utilizadas para medir la percepción sobre seguridad pública son:

- Sensación de inseguridad por temor a ser víctima de un delito: es la impresión general que tienen las personas sobre la situación de inseguridad en su entorno más próximo con base en su experiencia y lo vulnerable que se sienten de convertirse en víctimas del delito.
- Atestiguamiento de conductas antisociales y delictivas: identifica si los individuos han sido testigos o han escuchado si en su entorno más próximo hay presencia del delito o situaciones de desorden social; propiciando una opinión posiblemente desfavorable sobre la seguridad pública.
- Prevención y protección ante el delito: se identifican los cambios en las rutinas y vida cotidiana de los sujetos a consecuencia del temor a convertirse en víctimas del delito, al igual que medidas tomadas en los hogares y los costos derivados de estas acciones preventivas.
- Cohesión social y eficacia colectiva: se identifica la capacidad de los miembros de la comunidad para resolver los problemas que les afectan como conjunto y la forma en que llegan a acuerdos sobre sus propios controles y objetivos de bienestar social.

1.3 Contexto de la ENVIPE 2020

El INEGI presenta la Encuesta Nacional de Victimización y Percepción de la Seguridad Pública (ENVIPE) 2020. El propósito de esta encuesta es generar información a nivel nacional, por entidad federativa y áreas metropolitanas de interés sobre el fenómeno de la victimización delictiva durante 2019; así como la percepción social respecto de la seguridad pública y el desempeño de las autoridades en el periodo de marzo de 2020 para proveer información a la sociedad y a quienes toman decisiones de política pública en estas materias.

La ENVIPE permite hacer estimaciones estadísticas sobre la victimización que afecta de manera directa a las personas y a los hogares en los delitos de robo total de vehículo, robo parcial de vehículo, robo en casa habitación, robo o asalto en calle o transporte público. Asimismo, robo en forma distinta a las anteriores (como carterismo, allanamientos con robo en patio o cochera, abigeato), fraude, extorsión, amenazas verbales, lesiones y otros delitos distintos a los anteriores (como secuestros y delitos sexuales). Delitos como narcotráfico, delincuencia organizada, robo de combustible y trata de personas, entre otros en este tipo de categorías, no son susceptibles de medirse en una encuesta de victimización.

La Oficina de Naciones Unidas contra la Droga y el Delito (UNODC) identifica ocho actividades principales del crimen organizado: tráfico de drogas, trata de personas, tráfico ilícito de migrantes, tráfico ilícito de armas de fuego, tráfico ilícito de recursos naturales, comercio ilegal de flora y fauna, venta de medicamentos adulterados y delincuencia cibernética. Nuestro país no cuenta con un registro oficial que nos ayude a conocer cuántas y cuáles son las actividades o rubros de los grupos delictivos, pero de

acuerdo a una investigación¹, se puede hacer una aproximación de las distintas actividades del crimen organizado en el territorio nacional:

1. Trata de personas. Según datos del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP), durante 2020 se registraron 674 víctimas de trata de personas y 22 de tráfico de menores y hasta junio de 2021 se registraron 399 de trata de personas y 93 de tráfico de menores.³ Durante 2020, los grupos que realizaron dicha actividad fueron la Unión Tepito (activa en la Ciudad de México, Estado de México, Guerrero, Nuevo León, Veracruz y Oaxaca) y el Cártel Santa Rosa de Lima (CSRL), activo en Guanajuato. Ambos grupos incursionaron en la trata con fines de explotación sexual.
2. Tráfico de migrantes. El tráfico de migrantes puede tener la finalidad de facilitar el cruce ilegal de fronteras, pero también puede estar relacionado con la trata de personas y con la extorsión: los grupos criminales con frecuencia secuestran a las personas migrantes para someterlas a explotación sexual, a trabajo forzados o para extorsionarles. El Cártel de Sinaloa y grupos relacionados, como los Chapitos, operan en Sonora y Baja California respectivamente. En Tamaulipas, el Cártel del Golfo controla esta actividad.
3. Extorsión/cobro de piso. Según datos de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (Envipe), en 2020 se registraron 5,160 casos de extorsión por cada 100 000 personas. En los últimos años se ha registrado un aumento en la extorsión telefónica por parte de supuestos miembros de grupos criminales, cuya verdadera identidad no siempre es fácil de distinguir. En la Ciudad de México se han registrado llamadas de supuestos miembros del Cártel Jalisco Nueva Generación (CJNG), mientras que la Unión Tepito ha concentrado su actividad en el Centro Histórico, extorsionando a los comerciantes de la zona.
4. Secuestro. Se estima que durante 2020 hubo 80,954 víctimas de secuestros (Envipe, 2021). El secuestro, tanto económico como político, es una actividad común a muchos grupos del crimen organizado. Algunos ejemplos: el Cártel de Sinaloa lo realiza en Tijuana y hasta en Estados Unidos, mientras que el CJNG lo hace en Michoacán y los Beltrán Leyva en Nuevo León.
5. Piratería. Según el estudio Piratería en México, desarrollado en 2020 por el Observatorio Nacional Ciudadano, los principales mercados donde se vende piratería están en la Ciudad de México, el Estado de México, Guadalajara, Nuevo León, Puebla, Guanajuato y Chihuahua. Históricamente, el mercado de Tepito ha sido el centro de este tipo de actividades. Tal fue el origen de la Unión Tepito, organización que apareció entre 2009 y 2010 con la intención inicial de proteger a los comerciantes de Tepito y del Centro de robos y extorsiones. Actualmente, este grupo domina la piratería en México, seguido por el CJNG en Guadalajara y el Cártel del Golfo en Nuevo León.
6. Robo de combustible. Esta actividad, comúnmente conocida como «huachicol» en el caso de la gasolina y huachigas en el caso del gas, se ha convertido en otra de las principales fuentes de ingresos del crimen organizado. En 2020 se registraron robos de combustibles —gasolina, diésel y gas— equivalentes a un promedio de 5000 barriles al día. Según los datos más recientes de

¹ Pérez Dávila Samantha. Más allá del tráfico de drogas: la diversificación del crimen organizado. Noviembre 22, 2021. Consultada en <https://seguridad.nexos.com.mx/mas-alla-del-traffic-de-drogas-la-diversificacion-del-crimen-organizado/> el 02 de noviembre de 2022 a las 13:25 horas.

Petróleos Mexicanos (Pemex), las principales tomas clandestinas fueron detectadas en Guanajuato, Hidalgo, Jalisco, Estado de México, Puebla, Veracruz y Tamaulipas. Durante 2020, Pemex identificó como principales “huachicoleros” al Cártel del Noreste y el Cártel del Golfo en Tamaulipas, el CSRL en Guanajuato y al Grupo Sombra en Veracruz.

7. En 2020 se registraron 1121 denuncias por robo a vehículos en carreteras federales. Durante marzo y abril del mismo año, el robo a transporte de carga -en particular de alimentos y productos de primera necesidad- creció 7 % en el Estado de México, Veracruz, Puebla, Guanajuato, Jalisco, Michoacán, Hidalgo, Tlaxcala, Tabasco y la Ciudad de México. Si bien las autoridades no identificaron a ningún grupo en específico, un artículo de *Insight Crime* afirma que los principales carteles deben estar involucrados pues la mayoría de los robos se dan en rutas esenciales para el narcotráfico.
8. Tala ilegal. Entre 2018 y 2020 las autoridades atendieron 1176 denuncias por tala ilegal y se perdieron 127,770 hectáreas de bosque. Con base en notas periodísticas, sabemos que esta actividad se lleva a cabo en dos zonas: el Triángulo Dorado de Chihuahua, Sinaloa y Durango y la región entre Michoacán y Guerrero. Las autoridades han identificado a Los Viagras en Michoacán y en Guerrero y al grupo de “El Chano” Arreola en Petatlán, Guerrero como los principales responsables de esta actividad delictiva.
9. Extorsión a mineras. No hay un registro del número total de robos o ataques a empresas mineras, pero se sabe que se han llevado a cabo en el norte y sur del país en los últimos años. En Guerrero, se tiene registro de la extorsión del grupo Los Rojos a una empresa minera en 2011, la cual cerró años después debido a los niveles de violencia en el estado. Durante 2015, en Sinaloa se llevó a cabo un robo a una empresa canadiense. Y en 2020, un grupo sin identificar asaltó una minera en Sonora.
10. Tráfico de agua. En Morelos, el Comando Tlahuica amenazó a las autoridades para que estas le entregaran el control de la distribución de agua, argumentando que éste había sido entregado previamente a Los Rojos.
11. Tráfico de equipo médico y de medicamentos, incluyendo productos sin registro y falsificados. Según la UNOCD, ante el aumento de la demanda debido a la pandemia, se reportaron diversos casos de tráfico de equipo médico y de medicamentos robados, sin registro y falsificados en diversos países. En México, se detectó un amplio mercado ilegal de vacunas contra el covid-19. Según notas periodísticas, el CJNG es el principal grupo delictivo que trafica con fármacos ilegales en Guanajuato, Jalisco, Guerrero y Michoacán.
12. Ante la restricción de liquidez a la cual muchos negocios debieron enfrentarse, la Unión Tepito ofreció “minipréstamos” por un monto aproximado de 10,000 pesos con un interés del 5% en seis meses, según notas periodísticas.

Los resultados de análisis señalan que en México los grupos delictivos realizan al menos doce actividades ilegales -además del tráfico de drogas- en dieciocho entidades federativas. Si bien la mayoría de estas actividades han sido realizadas desde hace varios años, algunas de ellas se exacerbaron durante la pandemia para contrarrestar la pérdida de recursos de los grupos delictivos. Otras actividades, en cambio, aparecieron gracias a la pandemia, pues los grupos del crimen organizado aprovecharon los nichos de mercado que la crisis sanitaria y económica trajo consigo. Dentro de este último grupo se

encuentran el tráfico de equipo médico y medicamentos y la oferta de microcréditos. Tener un registro público mucho más detallado de todas estas actividades, quién las lleva a cabo y en qué zonas, permitirá entender mejor el rol de cada una de ellas dentro del financiamiento de los grupos delictivos. Esto resulta vital si se quieren diseñar mejores estrategias de combate al crimen organizado, que estén basadas en inteligencia local.

Dado su diseño conceptual y estadístico, la ENVIPE se ha consolidado como una de las mejores prácticas internacionales para medir victimización y delitos. El INEGI, en colaboración con la Oficina de Naciones Unidas contra la Droga y el Delito (UNODC) continúa compartiendo las experiencias de México en el diseño de encuestas de victimización tanto en hogares como en unidades económicas, para el desarrollo de encuestas nacionales y regionales.

Debido al problema que representa la denominada «cifra negra²» (subregistro de delitos) en los registros administrativos de delitos, las encuestas de victimización se constituyen en la alternativa para hacer la aproximación más apegada a la realidad del fenómeno de la delincuencia.

Las encuestas de victimización, a pesar de su enorme valor para proporcionar información sobre la victimización de personas, hogares o empresas del sector privado, no permiten medir delitos en los que no hay una víctima directa o donde la víctima no es una persona (delitos sin víctima identificable), tales como el lavado de dinero, el tráfico de drogas, de armas o de personas, entre otros. Así, las encuestas de victimización, como es el caso de la ENVIPE 2020, captan únicamente delitos que afectaron de manera directa a las víctimas y que ocurrieron durante 2019 a hogares y a personas de 18 años y más, integrantes de los hogares.

1.3.1 Implementación de políticas públicas

El diseño e implementación de la política pública en materia de prevención de la violencia y el delito debe ser efectiva y coordinada entre las dependencias y entidades de los tres órdenes de gobierno, sector privado, sociedad civil; así como organismos internacionales. El enfoque debe considerar las necesidades de la población.

Lo anterior basado en los resultados arrojados por la ENVIPE pues se han generado incrementos constantes en la prevalencia delictiva. Por ejemplo: según datos de la ENVIPE 2019, durante 2018, 33.9% de los hogares mexicanos tuvo al menos una víctima de delito, incluidos: robo total o parcial de vehículo; robo en casa habitación; robo o asalto en calle o transporte público; robo en general; fraude; extorsión; amenazas verbales; lesiones, secuestro y delitos sexuales, entre otros.

Como se indica previamente, los resultados de la ENVIPE 2019 señalan que alrededor de 24.7 millones de personas de 18 años o más han sido víctimas de delito, lo que representa una tasa de 28 mil 269 víctimas por cada 100 mil habitantes durante 2018, en una tendencia que muestra un comportamiento ascendente (véase cuadro 1.1).

² Se denomina Cifra negra o Cifra Oscura en criminología, al número de delitos y delincuentes que no han llegado a ser descubiertos o condenados. Es lo que constituye la cifra negra o cifra oscura de la criminalidad.

Cuadro 1.1
Tasa de víctimas de delito por cada 100 mil habitantes por año

Año	Tasa por cada 100 mil habitantes
2012	27,337
2013	28,224*
2014	28,200
2015	28,202
2016	28,788
2017	29,746*
2018	28,269*

*En estos casos existió un cambio estadísticamente significativo con relación al año anterior.

Fuente: ENVIPE 2019.

Por su parte, la ENVIPE 2019 muestra que la percepción de la población respecto a la situación actual sobre la inseguridad pública en su colonia, localidad, municipio o alcaldía y entidad federativa ha ido en aumento al comparar los niveles de 2013 a 2019, a la vez que se reflejan diferencias en la percepción de mujeres y hombres (véase cuadro 1.2).

Cuadro 1.2
Porcentaje de población que percibe inseguridad en el Municipio o Alcaldía

Año	Mujeres	Hombres	General
2013	65.6	60.0	63.0
2014	67.1*	60.8	64.2*
2015	67.3	60.5	64.1
2016	68.1	61.7*	65.1*
2017	69.5*	62.7*	66.3*
2018	73.4*	66.2*	70.0*
2019	74.1	66.3	70.5

*En estos casos existió un cambio estadísticamente significativo con relación al año anterior.

Fuente: ENVIPE 2019.

De conformidad con la propia ENVIPE 2019, la percepción de inseguridad en los ámbitos más próximos a las personas en el período de marzo y abril de 2019, fue de 70.5% y 50.6% de la población de 18 años y más que se siente insegura en su municipio o demarcación territorial y en su colonia o localidad, respectivamente.

1.3.2 Prácticas internacionales para la medición

Las Encuesta de Victimización, a nivel internacional, toman las directrices emitidas por la UNODC a través del Manual para Encuestas de Victimización, documento que sienta las bases de requerimientos mínimos para la generación e implementación de una encuesta nacional de victimización. Este Manual es el esfuerzo más consolidado para homologar las encuestas nacionales de victimización a nivel internacional; además en este se describen características básicas que deberían cumplir las encuestas de victimización mediante algunos atributos que deberían ser comparables entre las encuestas. Para efectos de dicha comparación, se toman en cuenta los siguientes parámetros:

Características metodológicas:

- Tipo de proyecto (Encuesta Nacional/Módulo).
- Tipo de entrevista.
- Periodicidad.
- Población objetivo.
- Tamaño de muestra.
- Porcentaje de respuesta positiva.

Características conceptuales:

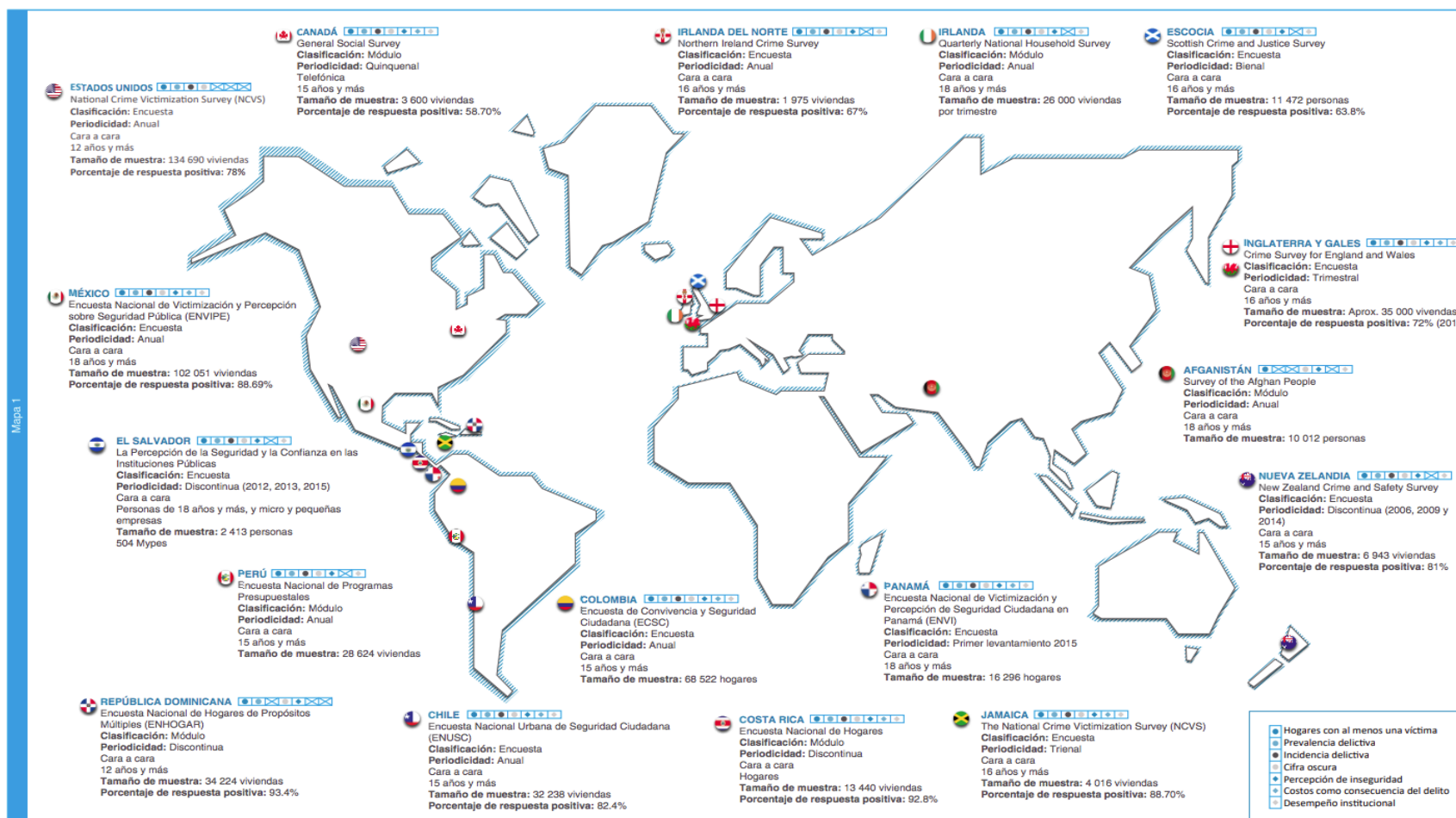
- Prevalencia delictiva (delitos contra el hogar y contra la persona).
- Incidencia delictiva.
- Cifra oscura o negras.
- Percepción de inseguridad.
- Desempeño institucional.

Para la selección de casos a comparar se siguieron los siguientes criterios:

- Última publicación al cierre de la investigación (diciembre 2017)
- Disponibilidad de la documentación publicada en español o inglés
- Última publicación de la Encuesta Nacional de 2010 a 2017
- Se excluyeron estudios, informes, compilado de varios países, programas regionales o mundiales; únicamente se consideraron encuestas nacionales a nivel internacional
- Se incluyeron Encuestas Nacionales que tuvieran un módulo o encuesta sobre victimización.

A continuación, se muestra un cuadro comparativo de Encuestas de Victimización: (Figura 1.1)

Figura 1.1
Cuadro comparativo de encuestas de victimización a nivel internacional



Fuente: Encuesta Nacional de Victimización a nivel internacional. UNODC. 2020

1.3.3 Captación de información durante la contingencia sanitaria (COVID-19)

Debido a la contingencia sanitaria generada a causa de la enfermedad por coronavirus (COVID-19), en atención a las disposiciones de las autoridades federales y locales, y a la determinación del INEGI de priorizar la salud de los entrevistadores y de la población informante, la etapa de captación de la ENVIPE 2020 prevista para los meses de marzo y abril de 2020 fue suspendida el 31 de marzo y se reanudó el 27 de julio. Por lo anterior, la captación de información se realizó en dos periodos del 17 al 31 de marzo y del 27 de julio al 04 de septiembre. En las ediciones anteriores de la ENVIPE el levantamiento de información se realizó sólo en los meses de marzo y abril.

Por esta circunstancia, los resultados relacionados con las experiencias de la población relativos a delitos podrían presentar algún grado de subestimación tanto en víctimas como en incidencia delictiva. Esto porque posiblemente los informantes no recordaron hechos relacionados con los actos delictivos experimentados durante 2019.

1.4 Diseño estadístico

El cuadro 1.3 muestra el diseño estadístico que se creó para esta encuesta:

Cuadro 1.3
Diseño estadístico

Periodo de referencia de la información	Enero-diciembre de 2019 para victimización Marzo de 2020 para percepción sobre seguridad pública y desempeño de las autoridades
Selección de la muestra	Probabilístico: trietápico, estratificado y por conglomerados
Unidades de observación	Las viviendas seleccionadas, los hogares, los residentes del hogar y las personas seleccionadas en el hogar
Población objeto de estudio	Población de 18 años y más
Tamaño de muestra nacional	101,988 viviendas
Periodo de levantamiento	17 al 31 de marzo y del 27 de julio al 04 de septiembre
Cobertura geográfica.	A nivel nacional y del 27 de julio al 04 de septiembre de 2020

Fuente: Elaboración propia a partir de ENVIPE 2020

1.4.1 Objetivos específicos

Los objetivos específicos que tiene son los siguientes:

- Hacer estimaciones sobre el número de personas que sufrieron victimización delictiva en 2019.
- Calcular el número de delitos ocurridos durante 2019.
- Hacer estimaciones sobre la «cifra negra» de los delitos y sus causas.
- Medir la percepción actual (2020) de la población de 18 años y más sobre la seguridad del lugar donde viven (colonia o localidad, municipio, entidad federativa) y sobre los lugares en donde realizan sus actividades cotidianas (trabajo, escuela, mercado, transporte público, entre otros).
- Medir el grado de confianza en las instituciones de seguridad pública y la percepción sobre el desempeño de las autoridades de seguridad pública y justicia.
- Identificar los cambios en las actividades de las personas por temor al delito.

- Estimar los gastos de los hogares para protegerse de la delincuencia.
- Conocer la repercusión del delito sobre las víctimas.
- Reunir información sobre las características sociodemográficas de las víctimas.
- Reunir información sobre las características del victimario y su relación con la víctima.
- Identificar actitudes y experiencias de las víctimas con las instituciones de seguridad pública y de procuración de justicia.

1.4.2 Diseño muestral

El diseño muestral puede entenderse como el conjunto de estrategias y procedimientos encaminados a seleccionar una muestra de una población objetivo de estudio, que cumple con una serie de características estadísticamente deseables que sirven para asegurar niveles establecidos de precisión y mantener en lo posible, la inclusión de errores y sesgos baja. Para la ENVIPE 2020 el diseño muestral es probabilístico: por conglomerados, estratificados, trietápicos.

El muestreo probabilístico es un método de muestreo que utiliza formas de métodos de selección aleatoria. El requisito más importante del muestreo probabilístico es que todos en una población tengan la misma oportunidad de ser seleccionados y ofrece la mejor oportunidad de crear una muestra representativa de la población. Este método utiliza la teoría estadística para seleccionar al azar un pequeño grupo de personas de una gran población existente y luego predecir que todas las respuestas juntas coincidirán con la población en general.

El muestreo por conglomerados y muestreo estratificado son técnicas de muestreo que tienen diferentes enfoques en cuanto a creación y análisis de las muestras.

El muestreo por conglomerados es un método en el que la población objetivo se divide en múltiples grupos. Algunos de estos grupos se seleccionan de manera aleatoria e incluso algunas veces, se lleva a cabo una segunda etapa o un muestreo de etapas múltiples para formar la muestra objetivo. Dependiendo del número de pasos seguidos para crear la muestra deseada, el muestreo de conglomerados se divide utilizando una sola etapa, dos etapas o técnicas de muestreo de múltiples etapas. Este método de muestreo es extremadamente rentable ya que requiere de un esfuerzo mínimo en cuanto a la creación de las muestras y también es conveniente para en cuanto a su ejecución.

El muestreo estratificado es un procedimiento en el que el objetivo de la población se separa en segmentos exclusivos, homogéneos (estratos), y luego una muestra aleatoria simple se selecciona de cada segmento (estrato). Los elementos de cada una de las muestras son distintos ya que por supuesto cada población tiene diferentes partes.

Finalmente, el muestreo polietápico, también conocido como muestreo multietapas, es un método que consiste en tomar muestras por etapas utilizando unidades de muestreo descendentes con el objetivo de hacer el proceso más práctico. Este método puede ser un poco complejo, ya que implica dividir la población en conglomerados y elegir uno o más grupos al azar y se muestrea a todos los miembros del grupo elegido.

1.4.3 Población objetivo

La población objetivo de la ENVIPE son los individuos de 18 años y más, que habitan las viviendas seleccionadas a nivel nacional y que sufrieron algún tipo de delito durante el 2019, hayan denunciado o no ante las autoridades dicho delito.

1.4.4 Levantamiento y cobertura

La cobertura de la ENVIPE es temporal ya que se consideró el 2019 para victimización y 2020 para percepción de la seguridad pública pues el periodo de levantamiento de información, debido a la emergencia sanitaria generada por el virus SARS-CoV2 (COVID-19), se realizó del 17 al 31 de marzo de 2019 y del 27 de julio al 04 de septiembre de 2020.

1.5 Análisis descriptivo

Una de las características metodológicas de la ENVIPE que la distingue es su periodicidad anual, lo que contribuye a tener información precisa y actualizada sobre la situación de inseguridad en México. La ENVIPE cuenta con un tamaño de muestra de 101,988 viviendas, lo que la posiciona como la segunda encuesta con la muestra más grande después de la Encuesta Nacional de Ingresos y Gastos en los Hogares (ENIGH) que tiene una muestra de 105,483 viviendas. La cifra negra es el principal objetivo de las encuestas de victimización; sin embargo, en México además de investigar los delitos no denunciados, también se exploran aquellos que no tuvieron un seguimiento por parte de la autoridad competente.

En México, desde el sexenio 2012-2018, el Plan Nacional de Desarrollo, albergó los ejes estratégicos para la generación de políticas públicas para la prevención y combate a la delincuencia, para ello, retomó la información estadística generada por la ENVIPE que permitió la elaboración del diagnóstico en materia de seguridad pública, además de la tasa de victimización como indicador para la construcción de la primera meta nacional: Un México en Paz³.

En los gastos a consecuencia de la inseguridad y el delito, la encuesta de victimización en México incluye: las erogaciones resultado de las medidas de prevención que adopta la gente para evitar ser víctima de un delito; gastos a consecuencia del delito, y gastos en salud a consecuencia del delito. La ENVIPE en México se considera como información de interés nacional, lo que significa que sus resultados son de carácter oficial y de uso obligatorio para la Federación, los estados y los municipios. Por ello, debe ser usada para la toma de decisiones en el proceso de políticas públicas en la materia.

La ENVIPE se ha convertido en uno de los referentes para dar seguimiento a los avances en el cumplimiento de los objetivos de desarrollo sustentable, debido a que genera información sobre la prevalencia de la población sometida a violencia física, psicológica o sexual en los últimos doce meses; además estima la proporción de personas víctimas de un delito que reportaron a las autoridades correspondientes.

El INEGI genera y publica documentos de cada una de las etapas de la encuesta (informe operativo, diseño muestral y marco conceptual); al igual que los instrumentos de captación (cuestionario principal y módulo sobre victimización) utilizados para el levantamiento. Además, pone a la disposición del

³6 líneas de acción para un México en Paz. <https://www.gob.mx/epr/es/articulos/6-lineas-de-accion-para-un-mexico-en-paz> consultado el 02 de noviembre de 2022 a las 14:30 horas

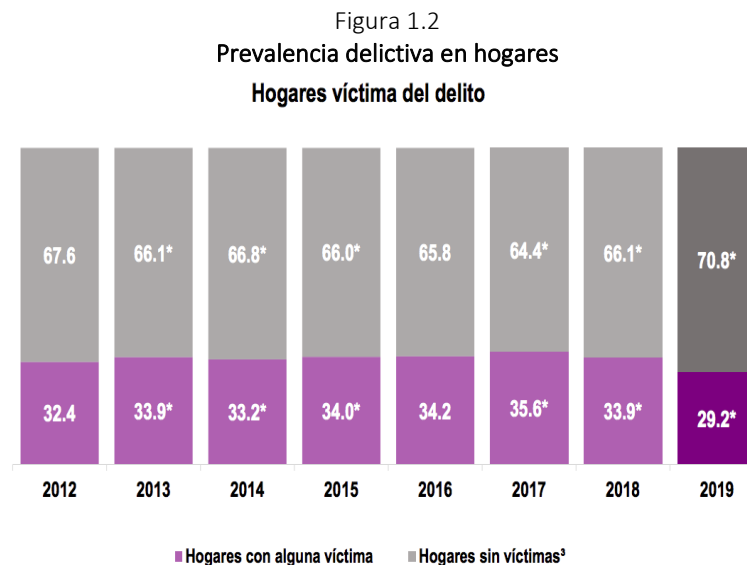
usuario el glosario de términos; los tabulados que incluyen los resultados de cada una de las variables; una presentación de los principales datos a nivel nacional y por entidad federativa; así como los microdatos que contienen todos los datos resultantes de la aplicación de la encuesta y el archivo “descriptor” de la base de datos que contiene las características de los microdatos. Estos documentos son indispensables para aquellos que consultan la encuesta, quienes pueden tener la certeza de la confiabilidad en cada uno de los procesos de la encuesta.

La difusión de los resultados obtenidos en cada una de las etapas que conforman la ENVIPE es otro punto para resaltar: se realiza una conferencia de prensa donde asisten los principales medios de comunicación de televisión, radio y periódicos, que retoman la información para la generación de notas informativas y reportajes. Además, en esta estrategia también intervienen las redes sociales (Facebook, Twitter) del Instituto y del Centro de Excelencia, lo que le da una proyección internacional a la encuesta.

En conclusión, se puede afirmar que la encuesta de victimización en México es una buena práctica a nivel internacional, ya que responde a las necesidades de información tanto del público nacional como internacional, así como un referente para la instrumentación de políticas gubernamentales en materia de seguridad y prevención del delito. Adicionalmente, la ENVIPE se encuentra constantemente en revisión con el objetivo de cubrir las nuevas necesidades de información mediante el apoyo técnico de la Oficina de las Naciones Unidas contra la Droga y el Delito, a través del Centro de Excelencia para Información Estadística de Gobierno, Seguridad Pública, Victimización y Justicia.

1.5.1 Prevalencia delictiva en hogares

La prevalencia es la proporción de objetivos del delito que fueron de facto victimizados, es decir, es un cociente del número de víctimas entre el número total de objetivos potenciales de victimización. Durante 2019, el 29.2% de los hogares del país tuvo, al menos, una víctima de delito (Figura 1.2)

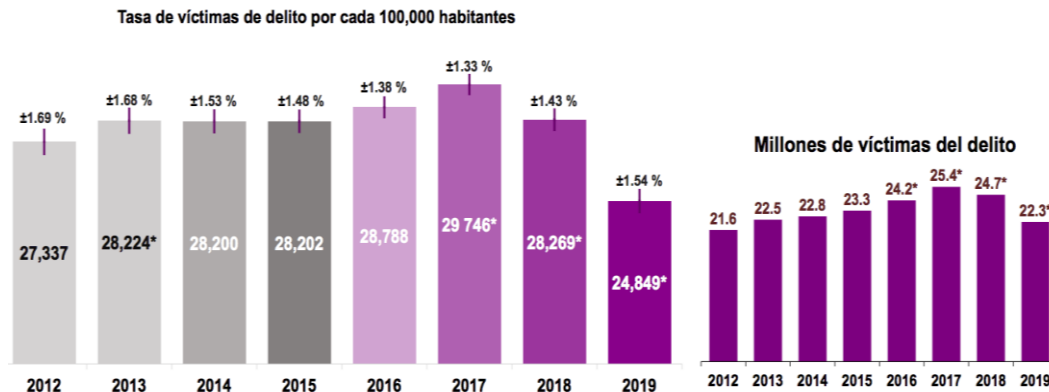


Fuente: Elaboración propia a partir de ENVIPE 2020

1.5.2 Prevalencia delictiva en las personas

A nivel nacional, se estiman 22.3 millones de víctimas de 18 años y más, lo cual representa una tasa de 24,849 víctimas por cada cien mil habitantes durante 2019 (Figura 1.3)

Figura 1.3
Prevalencia delictiva en las personas

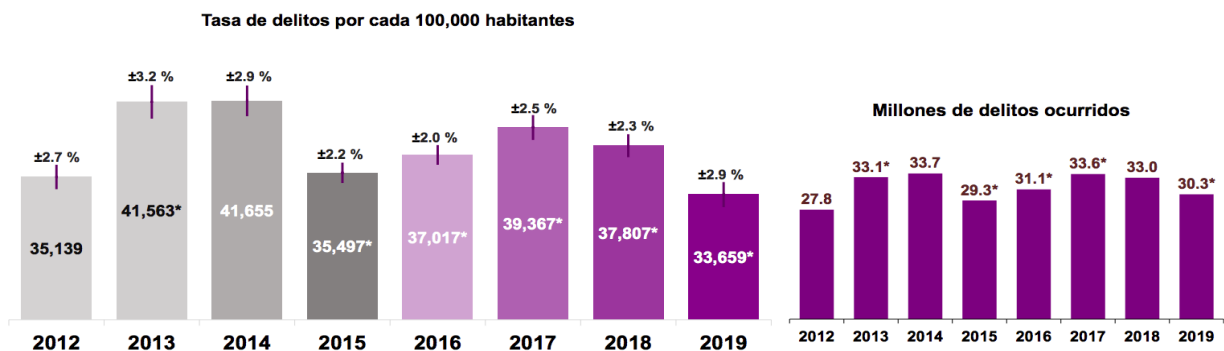


Fuente: Elaboración propia a partir de ENVIPE 2020

1.5.3 Incidencia delictiva

La incidencia delictiva se refiere a la presunta ocurrencia de delitos registrados en averiguaciones previas iniciadas o carpetas de investigación, reportadas por las Procuradurías de Justicia y Fiscalías Generales de las entidades federativas en el caso del fuero común y por la Fiscalía General de la República en el fuero federal (ENVIPE 2020). A nivel nacional, se estima una tasa de 33,659 delitos por cada cien mil habitantes durante 2019. Las pruebas de hipótesis demuestran que, estadísticamente, la incidencia delictiva tuvo un descenso respecto al año anterior (Figura 1.4)

Figura 1.4
Incidencia delictiva en las personas



Fuente: ENVIPE 2020

En el cuadro 1.4 se muestra la tasa de delitos por cada cien mil habitantes para la población de 18 años y más en 2019, según la entidad federativa en que ocurrió el delito (Cuadro 1.4)

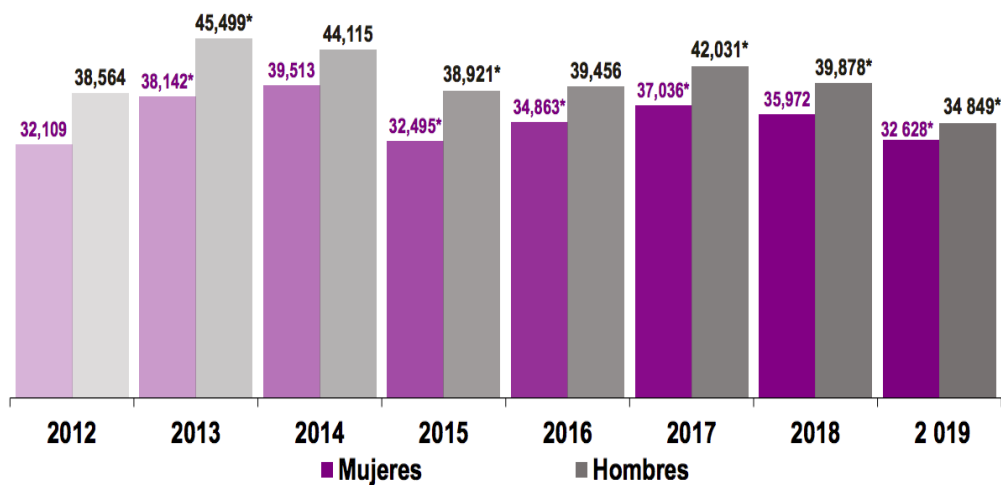
Cuadro 1.4
Incidencia delictiva 2018-2019
(Tasa de delitos)

Entidad	Delitos 2018	Delitos 2019	Cambio (Δ %)	Entidad	Delitos 2018	Delitos 2019	Cambio (Δ %)	Entidades Federativas
NACIONAL	37 807	33 659	-11.0*					A la baja ↓ 17
Aguascalientes	36 500	48 443	32.7*	Morelos	45 312	40 298	-11.1	A la alza ↑ 1
Baja California	42 725	35 655	-16.5*	Nayarit	23 670	18 214	-23.1*	Sin cambio ↔ 14
Baja California Sur	28 377	18 887	-33.4*	Nuevo León	27 805	28 319	1.8	
Campeche	26 466	18 595	-29.7*	Oaxaca	26 221	22 069	-15.8	
Coahuila	24 813	21 896	-11.8	Puebla	37 647	33 014	-12.3	
Colima	28 376	29 280	3.2	Querétaro	32 756	36 034	10.0	
Chiapas	19 409	13 903	-28.4*	Quintana Roo	33 243	35 535	6.9	
Chihuahua	28 622	23 133	-19.2*	San Luis Potosí	32 342	29 384	-9.1	
Ciudad de México	69 716	62 008	-11.1*	Sinaloa	29 507	25 945	-12.1*	
Durango	22 586	21 373	-5.4	Sonora	50 861	25 646	-49.6*	
Guanajuato	38 067	50 894	33.7	Tabasco	36 546	32 148	-12.0*	
Guerrero	43 051	30 864	-28.3*	Tamaulipas	25 368	21 954	-13.5*	
Hidalgo	25 987	20 732	-20.2*	Tlaxcala	40 336	30 177	-25.2*	
Jalisco	40 543	34 703	-14.4*	Veracruz	25 350	19 542	-22.9*	
Estado de México	51 520	48 916	-5.1	Yucatán	26 462	17 686	-33.2*	
Michoacán	22 999	20 414	-11.2	Zacatecas	26 670	22 363	-16.2	

Fuente: ENVIPE 2020

Históricamente, los hombres son víctimas de un mayor número de delitos que las mujeres. Durante 2019, la incidencia delictiva en hombres fue 7% mayor que la estimada para mujeres (Figura 1.5)

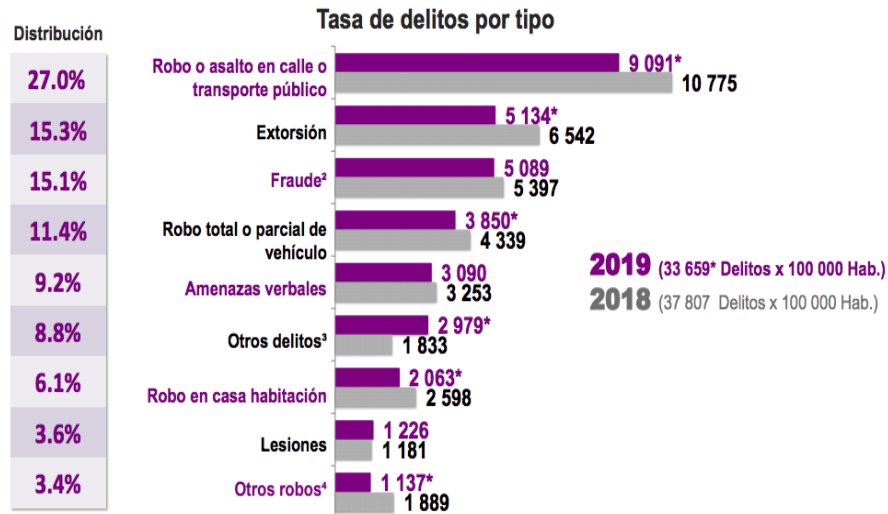
Figura 1.5
Incidencia delictiva en las personas por sexo
Tasa de delitos por sexo



Fuente: ENVIPE 2020

Durante 2019 se cometieron 30.3 millones de delitos asociados a 22.3 millones de víctimas. Esto representa una tasa de concentración de 1.4 delitos por víctima; en 2018 fue de 1.3 (Figura 1.6)

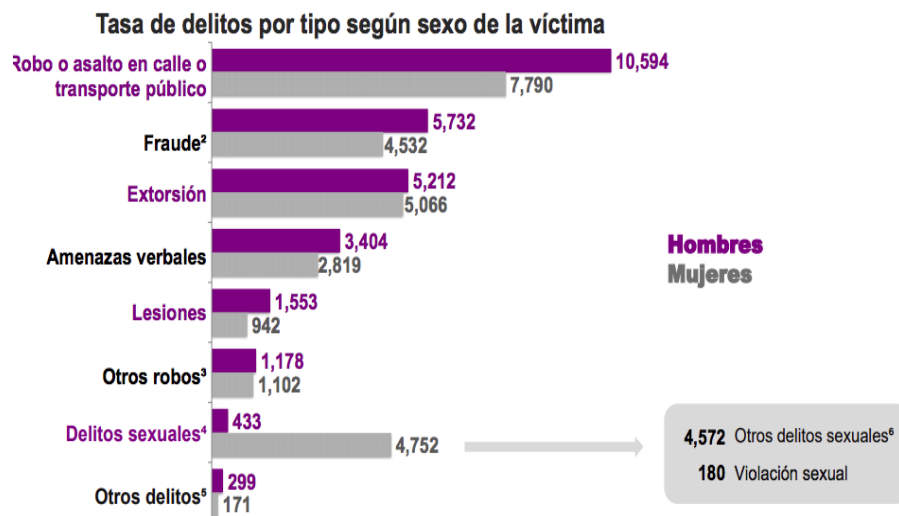
Figura 1.6
Incidencia por tipo de delito



Fuente: ENVIPE 2020

La incidencia delictiva es mayor en los hombres para la mayoría de los delitos. Sin embargo, en los delitos sexuales las mujeres se ven más vulneradas al contabilizarse 9 delitos sexuales cometidos a mujeres por cada delito sexual cometido a los hombres (Figura 1.7)

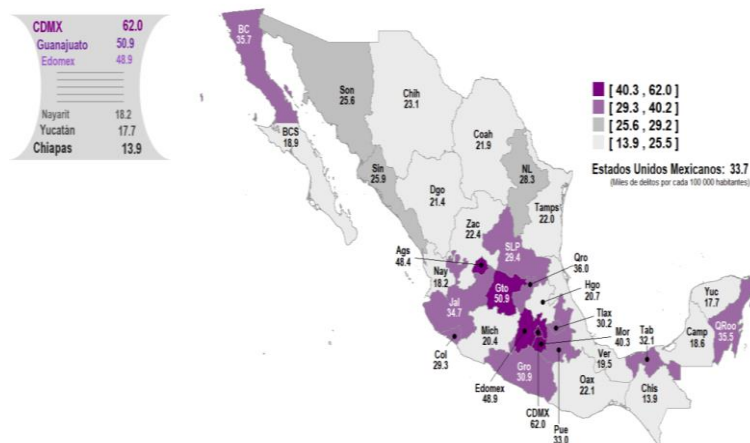
Figura 1.7
Incidencia delictiva por sexo



Fuente: ENVIPE 2020

La tasa de delitos por cada cien mil habitantes para la población de 18 años y más en 2019, por entidad federativa de ocurrencia del delito se presenta en la figura 1.8

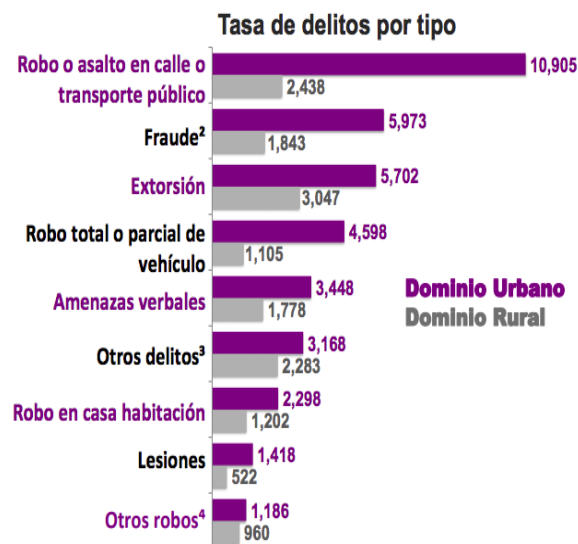
Figura 1.8
Incidencia delictiva por entidad federativa



Fuente: ENVIPE 2020

Durante 2019, se estima un total de 27.3 millones de delitos asociados al dominio urbano. Esto representa una tasa de 28.696 delitos por cada 100 mil habitantes, mientras que en el dominio rural se estima un total de 2.9 millones de delitos, con una tasa de 15,178 delitos por cada 100 mil habitantes (Figura 1.9)

Figura 1.9
Incidencia delictiva en las personas



Fuente: ENVIPE 2020

En el cuadro 1.5 se muestra la tasa de delitos por cada cien mil habitantes para la población de 18 años y más en 2019, según entidad federativa en que ocurrió el delito (Cuadro 1.5)

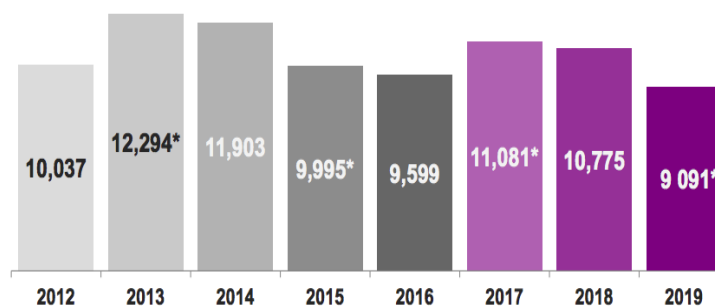
Cuadro 1.5
Incidencia delictiva según dominio 2018-2019
 (Tasa de delitos)

Entidad	Delitos 2018		Delitos 2019		Entidad	Delitos 2018		Delitos 2019	
	Dominio Urbano	Dominio Rural	Dominio Urbano	Dominio Rural		Dominio Urbano	Dominio Rural	Dominio Urbano	Dominio Rural
NACIONAL	43 333	17 350	38 696*	15 178					
Aguascalientes	39 625	21 237	53 722*	23 392	Morelos	46 144	40 686	43 197	24 317*
Baja California	43 239	35 352	35 977*	31 695	Nayarit	28 692	12 054	22 979	NS
Baja California Sur	30 566	14 297	19 169*	NS	Nuevo León	28 470	14 447	29 110	12 875
Campeche	31 471	10 443	21 632*	8 918	Oaxaca	38 263	14 276	30 386	13 948
Coahuila	25 513	NS	23 067	11 040	Puebla	44 566	18 388	37 635*	20 504
Colima	28 060	NS	31 766	9 387	Querétaro	37 083	21 000	42 526*	17 913
Chiapas	27 861	10 208	20 836*	6 316*	Quintana Roo	35 305	16 052	38 680	9 018*
Chihuahua	31 832	9 354	25 629*	8 905	San Luis Potosí	41 101	15 550	40 663	NS
Ciudad de México	68 931	209 559	61 506*	NS	Sinaloa	32 242	21 775	30 010	14 685*
Durango	28 588	9 017	27 070	8 422	Sonora	55 059	24 921	28 673*	5 624*
Guanajuato	48 191	12 068	57 547	NS	Tabasco	42 414	27 828	38 561	22 515
Guerrero	56 640	22 669	39 644*	17 523	Tamaulipas	27 158	NS	24 437	NS
Hidalgo	36 626	12 327	27 628*	11 957	Tlaxcala	41 542	34 709	32 076*	20 916*
Jalisco	43 695	18 774	36 975*	19 198	Veracruz	34 418	9 939	26 101*	8 485
Estado de México	54 439	29 739	52 346	23 040	Yucatán	29 848	8 064	19 714*	5 643
Michoacán	25 602	16 942	24 343	11 333	Zacatecas	36 230	12 179	31 291	8 583

Nota: El orden de las entidades federativas es alfabético y no representa una lógica de análisis.
 Fuente: ENVIPE 2020

Durante 2019 se cometieron 8.2 millones de robos o asaltos en la calle o en el transporte público, lo cual representa una tasa de 9,091 robos por cada 100,000 habitantes. En 65.6% de los casos, lo robado fue dinero, tarjetas de crédito o cheques, mientras que en 56% de los casos fueron teléfonos celulares (Figura 1.10)

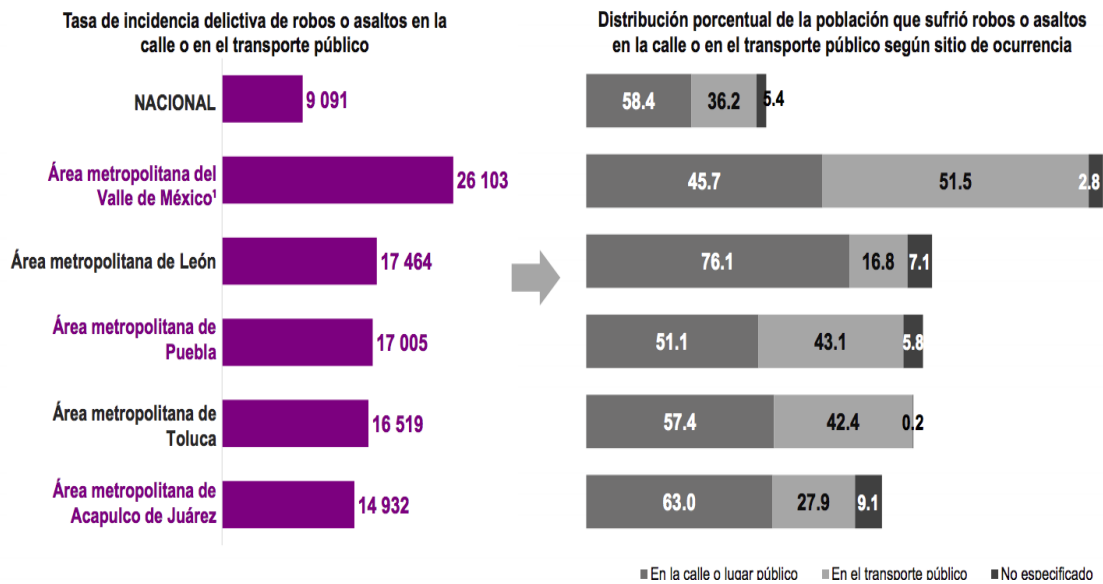
Figura 1.10
Incidencia delictiva por robo o asalto en la calle o en el transporte público
Tasa de delitos de robo en la calle o en el transporte público por cada 100 000 habitantes



Fuente: ENVIPE 2020

Durante 2019 en el Área Metropolitana del Valle de México se estima una tasa de 26,103 robos o asaltos en la calle o en el transporte público por cada 100,000 habitantes (Figura 1.11)

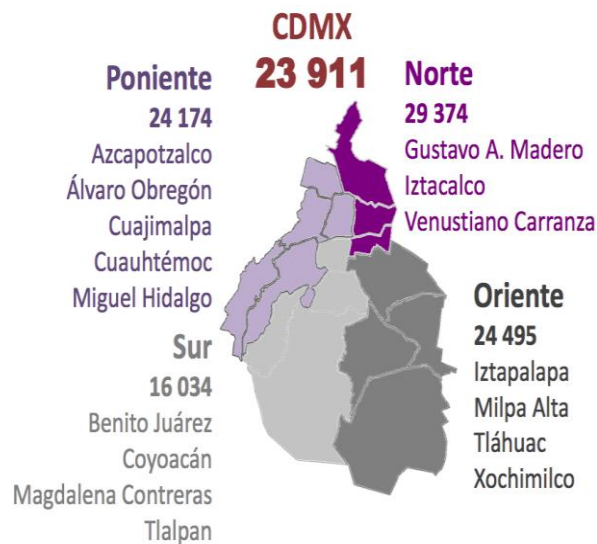
Figura 1.11
 Incidencia delictiva por robo o asalto en la calle o transporte según sitio de ocurrencia



Fuente: ENVIPE 2020

Durante 2019 en la Ciudad de México se estima una tasa de 23,911 robos o asaltos en la calle o en el transporte público por cada 100,000 habitantes (Figura 1.12)

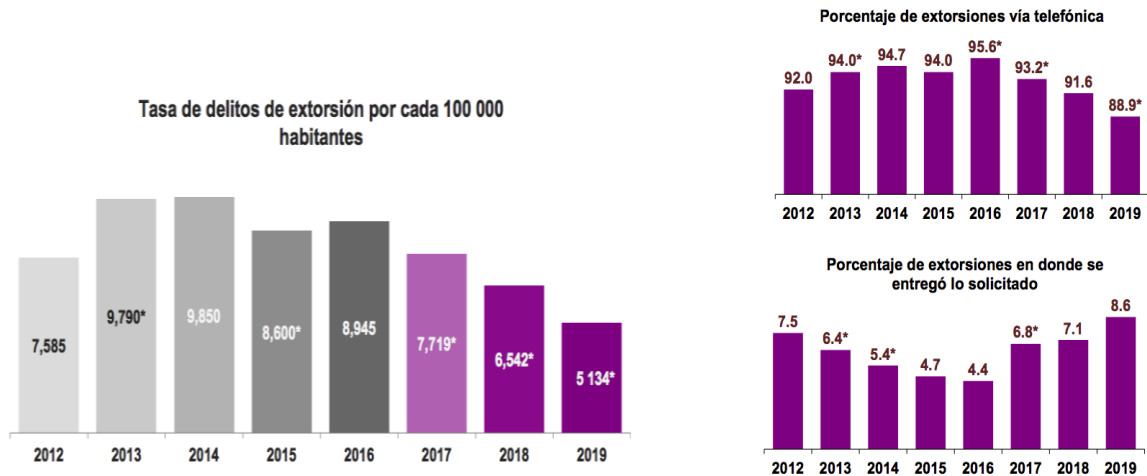
Figura 1.12
 Incidencia delictiva por robo o asalto en calle o transporte público en la CDMX



Fuente: ENVIPE 2020

Durante 2019 se cometieron 4.6 millones de delitos de extorsión, lo cual representa una tasa de 5,134 extorsiones por cada 100,000 habitantes. El 88.9% de los casos, la extorsión fue telefónica. La extorsión fue pagada en el 8.6% de las ocasiones (Figura 1.113)

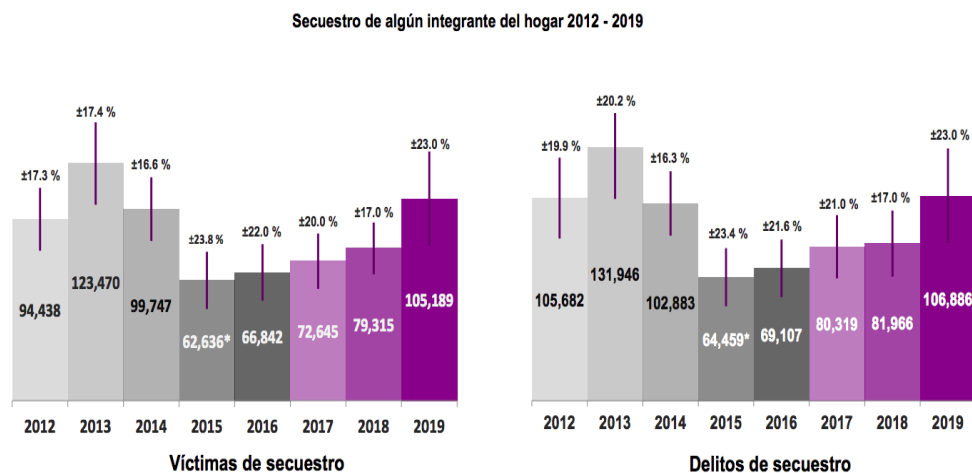
Figura 1.13
Incidencia delictiva por extorsión



Fuente: ENVIPE 2020

A nivel nacional, se estiman 106,886 secuestros de algún integrante del hogar sufridos por 105,189 víctimas durante 2019 (Figura 1.14)

Figura 1.14
Incidencia delictiva por secuestro de algún integrante del hogar



Fuente: ENVIPE 2020

Durante 2019m 63.7% de los secuestros estimados tuvieron una duración de menos de 24 horas (Cuadro 1.6)

Cuadro 1.6
Incidencia delictiva por secuestro de algún integrante del hogar.

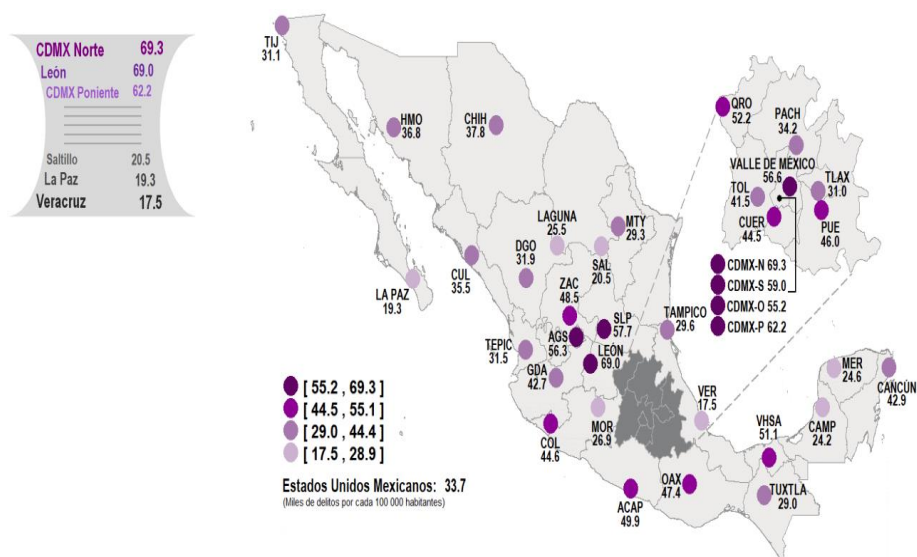
Secuestros	Tasas por cada 100 000 habitantes*							
	2012	2013	2014	2015	2016	2017	2018	2019
Víctimas (Prevalencia)	80	103	83	51*	54	58	63	83
Delitos (Incidencia)	89	110	85	53*	56	65	65	84

Duración de los secuestros	Distribución Porcentual							
	2012	2013	2014	2015	2016	2017	2018	2019
Menos de 24 horas	57.0	64.8	59.8	61.8	66.4	58.4	62.5	63.7
De 1 a 3 días	18.9	17.4	23.2	28.4	20.5	14.4	18.9	12.3
4 días o más	18.6	13.2	14.8	9.1^b	12.8^b	18.8	6.0*	10.9
No especificado	5.5	4.6^b	2.2^b	0.7^b	0.3*^b	8.3	12.5	13.1

Fuente: ENVIPE 2020

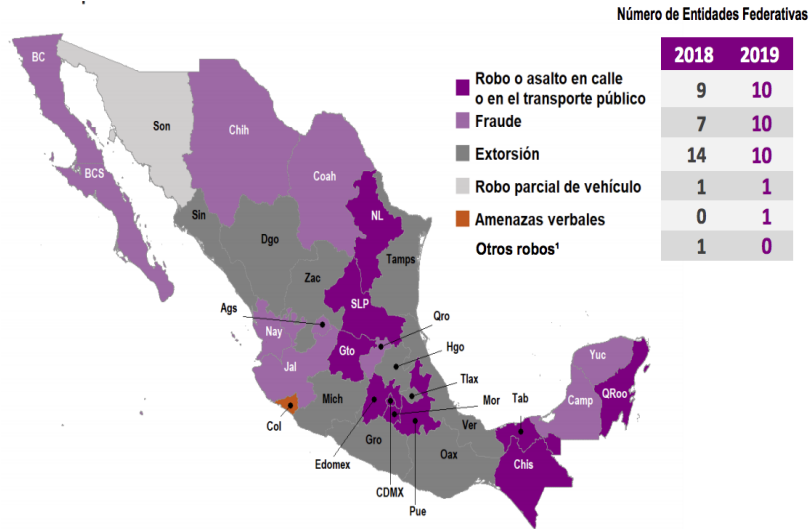
La tasa de delitos por cada cien mil habitantes para la población de 18 años y más por ciudad o área metropolitana de interés se muestra en la figura 1.15 y la incidencia de delitos más frecuentes en la figura 1.16, para esa misma población

Figura 1.15
Incidencia delictiva en las personas



Fuente: ENVIPE 2020

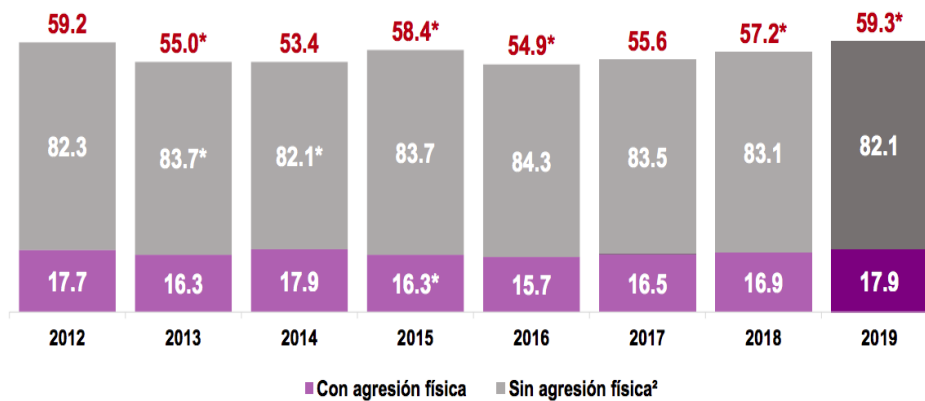
Figura 1.16
Incidencia de delitos más frecuentes



Fuente: ENVIPE 2020

De los 30.3 millones de delitos estimados para 2019, la víctima estuvo presente en 59.3% de los casos, lo que representa en términos absolutos 17.9 millones. De estas víctimas, el 17.9% involucró algún tipo de agresión física (Figura 1.17)

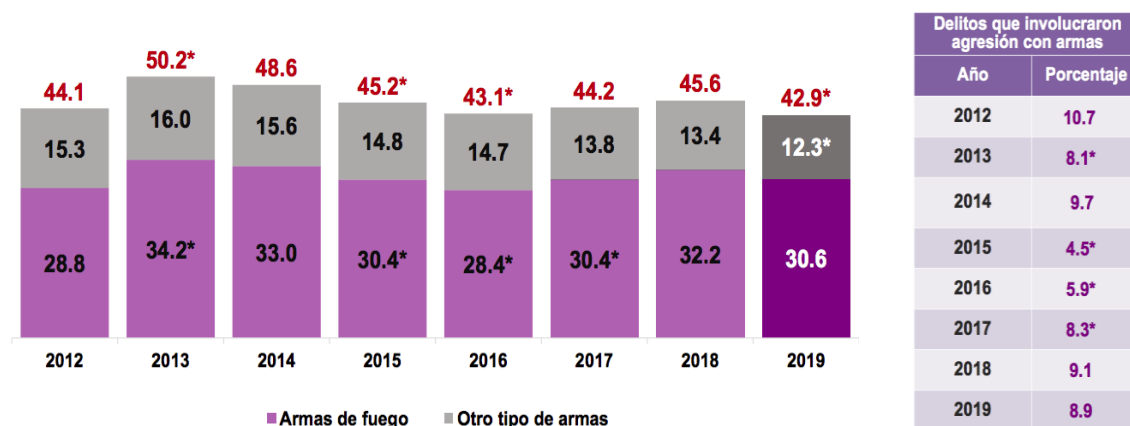
Figura 1.17
Porcentaje de delitos donde la víctima estuvo presente



Fuente: ENVIPE 2020

De los 17.9 millones de delitos estimados donde la víctima estuvo presente, los delincuentes portaban algún arma en 42.9% de los casos (Figura 1.18)

Figura 1.18
Porcentaje de delitos cometidos con portación de arma



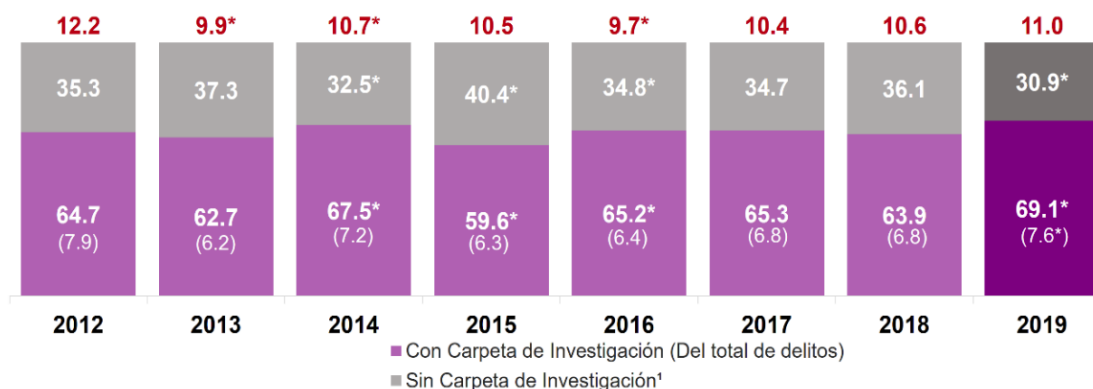
Fuente: ENVIPE 2020

1.5.4 Cifra Negra

La cifra negra hace referencia a aquellos datos que no son registrados en las estadísticas oficiales que no son revelados por las entidades policiales y judiciales, es decir, los delitos no denunciados o que no iniciaron una investigación (ENVIPE 2020)

La ENVIPE, estima que en 2019 se denunció el 11% del total de los delitos, y para los cuales el Ministerio Público inició una Carpeta de Investigación en 69.1% de los casos. Durante 2019 se inició Carpeta de Investigación en 7.6% del total de los delitos. Lo anterior, representa 92.4% de delitos donde no hubo denuncia o no se inició una Carpeta de Investigación (Figura 1.19)

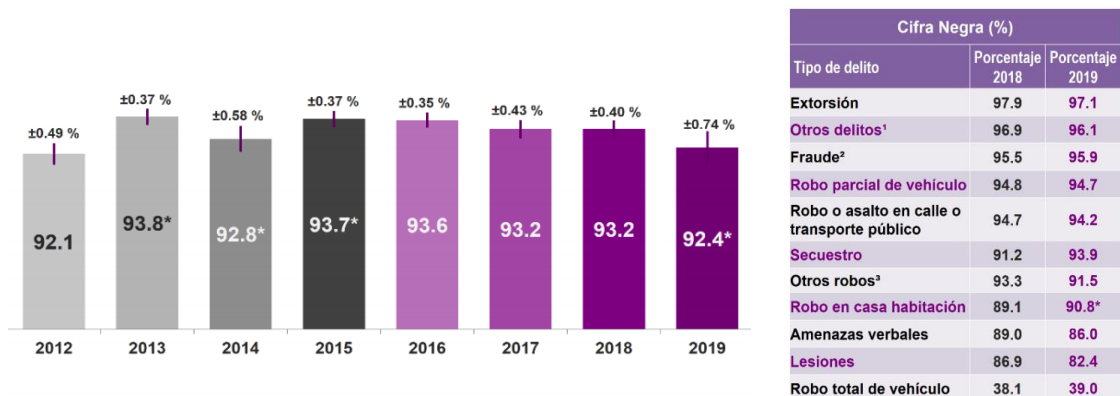
Figura 1.19
Porcentaje de delitos denunciados en el Ministerio Público



Fuente: ENVIPE 2020

En este sentido, la ENVIPE permite estimar que la “Cifra Negra” a nivel nacional en 2019 fue de 92.4%, cifra estadísticamente inferior a la estimación de 93.2% en 2018 (Figura 1.20)

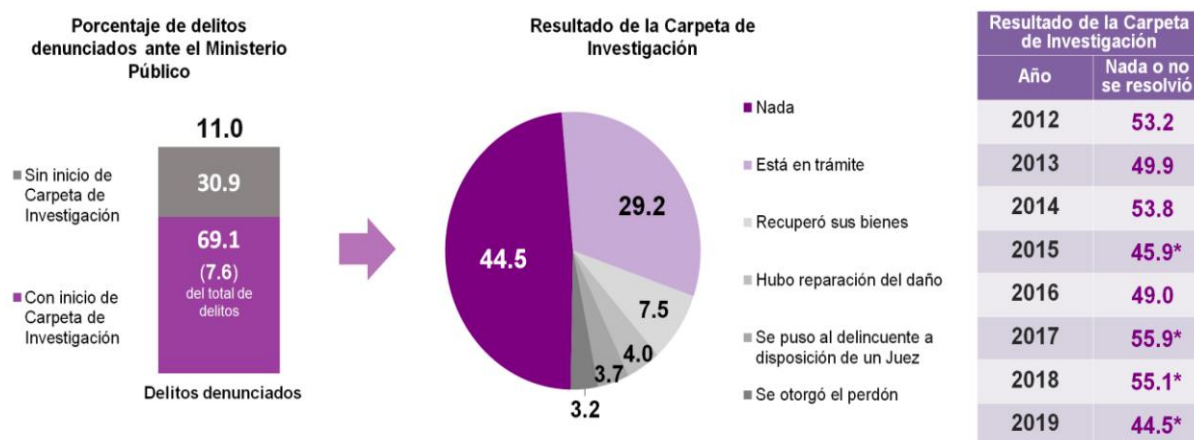
Figura 1.20
Cifra Negra



Fuente: ENVIPE 2020

Del total de Carpetas de Investigación iniciadas por el Ministerio Público, en el 44.5% de los casos no pasó nada o no se resolvió la denuncia (Figura 1.21)

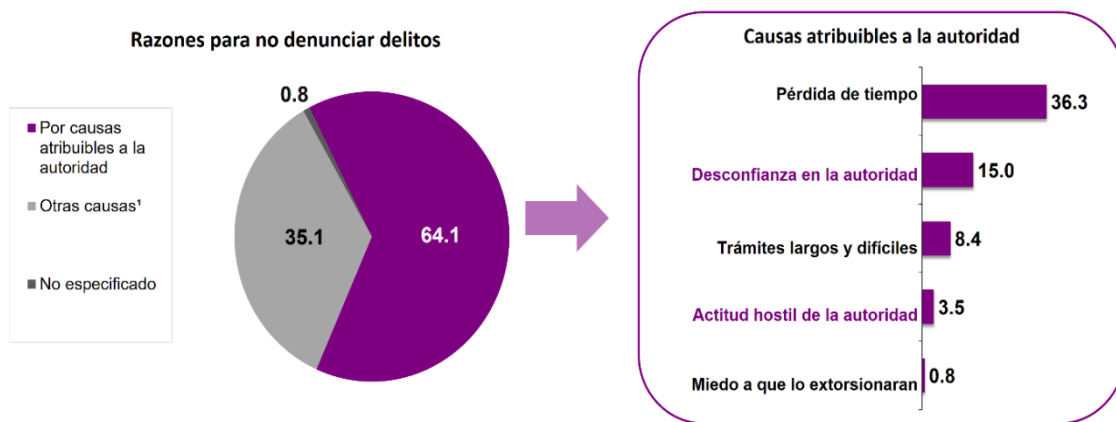
Figura 1.21
Resultados de la carpeta de investigación



Fuente: ENVIPE 2020

Entre las razones para no denunciar delitos ante las autoridades por parte de las víctimas destaca la pérdida de tiempo con 36.3% y la desconfianza en la autoridad con 15%, dentro de las causas atribuibles a la autoridad (Figura 1.22)

Figura 1.22
Tazones para no denunciar delitos

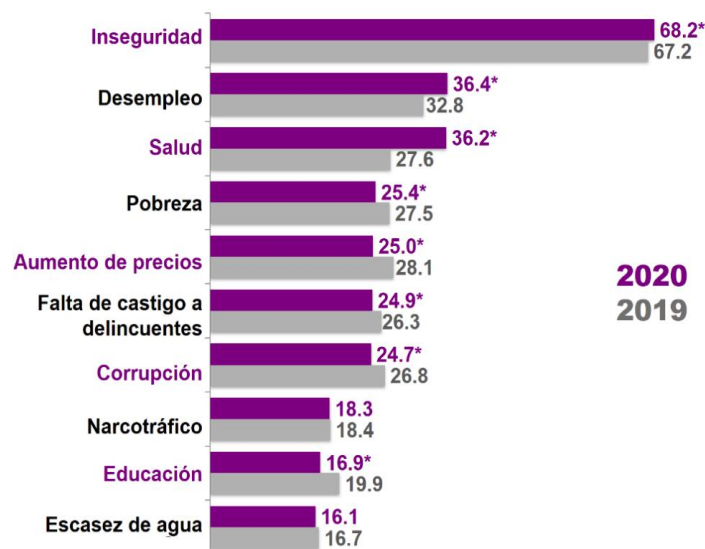


Fuente: ENVIPE 2020

1.5.5 Percepción sobre seguridad pública

La ENVIPE estima a nivel nacional que 68.2% de la población de 18 años y más considera la inseguridad como el problema más importante que aqueja hoy en día, seguido del desempleo con 36.4% y la salud con 36.2 % (Figura 1.23)

Figura 1.23
Distribución porcentual sobre los temas que generan mayor preocupación



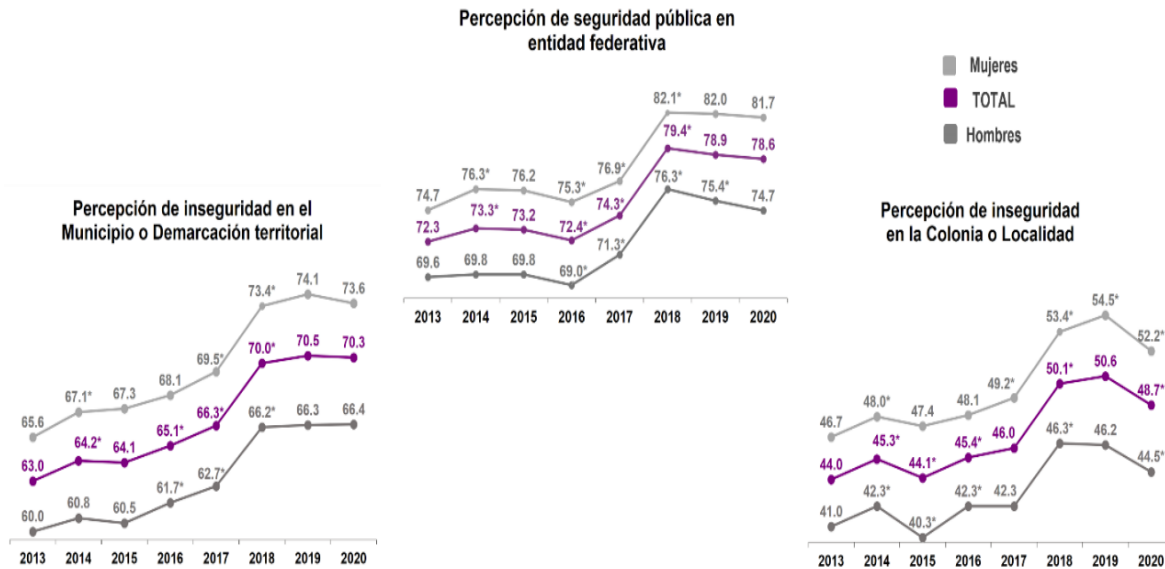
Los porcentajes no suman 100% porque se pueden tener más de dos opciones de respuesta.

Fuente: ENVIPE 2020

La ENVIPE estima que 78.6% de la población de 18 años y más considera que vivir en su entidad federativa es inseguro a consecuencia de la delincuencia. La sensación de inseguridad en los ámbitos más próximos a las personas disminuyó estadísticamente, con 70.3% y 48.7% de la población de 18

años y más que se siente insegura en su municipio o demarcación territorial y en su colonia o localidad, respectivamente (Figura 1.24)

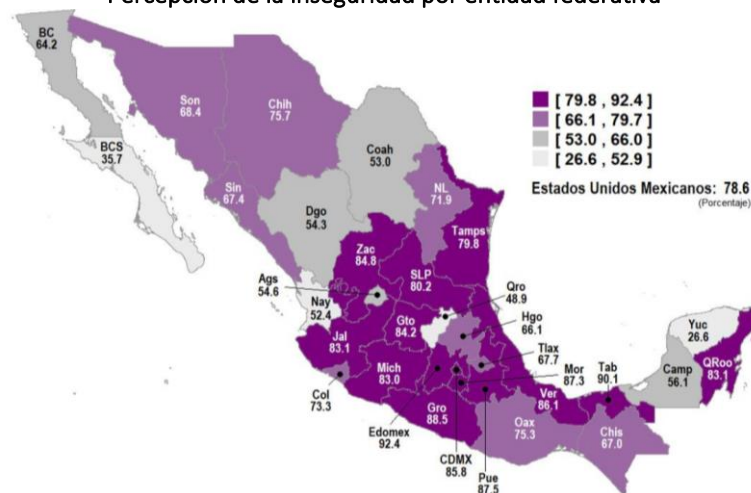
Figura 1.24
Percepción de la inseguridad y de la seguridad pública



Fuente: ENVIPE 2020

En el siguiente mapa se señala la percepción de la población respecto de la situación que guarda la inseguridad pública en su entidad federativa (Figura 1.25)

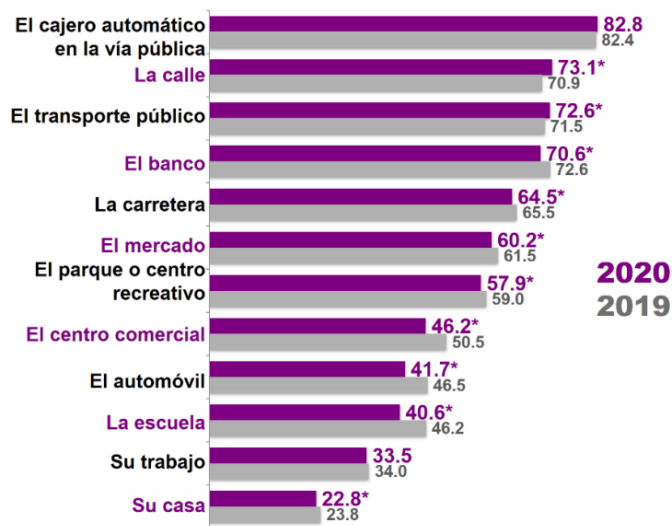
Figura 1.25
Percepción de la inseguridad por entidad federativa



Fuente: ENVIPE 2020

A nivel nacional, el espacio donde la población de 18 años y más se siente más insegura es el cajero automático ubicado en la vía pública con 82.8 por ciento (Figura 1.26)

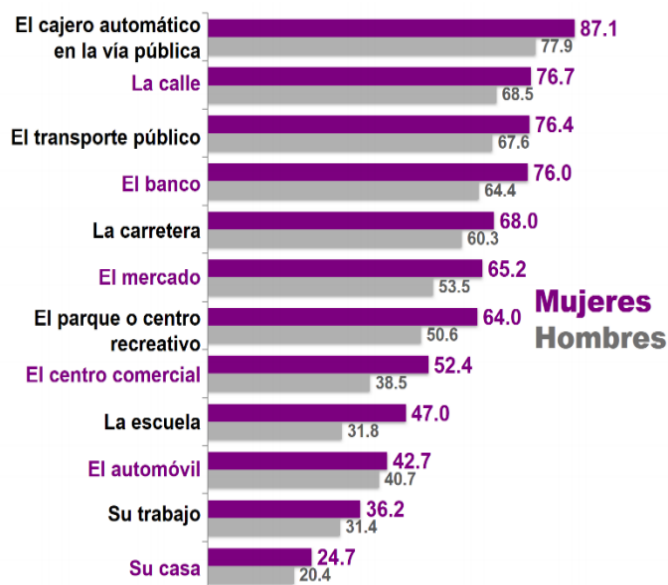
Figura 1.26
Distribución porcentual de la población que manifiesta sentirse insegura en espacios públicos o privados



Fuente: ENVIPE 2020

En términos generales, las mujeres se sienten más inseguras que los hombres en diversos lugares públicos o privados (Figura 1.27)

Figura 1.27
Distribución porcentual de la población que manifiesta sentirse insegura en espacios públicos o privados por sexo.



Fuente: ENVIPE 2020

1.6 Conclusión del capítulo

En el capítulo 1 se destacaron las características metodológicas de la ENVIPE 2020 entre las que destaca contribuir con información precisa y actualizada sobre la situación de inseguridad en México. La ENVIPE 2020 en México se considera como información de interés nacional, lo que significa que sus resultados son de carácter oficial y de uso obligatorio para Federación, los estados y los municipios. Por ello, debe ser usada para la toma de decisiones en el proceso de políticas públicas en la materia.

La ENVIPE se ha convertido en un referente, ya que genera información sobre la prevalencia de población sometida a violencia física, psicológica o sexual en los últimos doce meses; además estima la proporción de personas víctimas de un delito que reportaron a las autoridades correspondientes. Dichos temas se alinean particularmente con el indicador de «Porcentaje de la población sometida a violencia física, psicológica o sexual en los últimos 12 meses» y el indicador de «Porcentaje de las víctimas de violencia en los 12 meses anteriores que notificaron su victimización a las autoridades competentes u otros mecanismos de resolución de conflictos reconocidos oficialmente»

La difusión de los resultados obtenidos en cada una de las etapas que conforman la ENVIPE es otro punto para resaltar: se realiza una conferencia de prensa donde asisten los principales medios de comunicación de televisión, radio y periódicos, que retoman la información para la generación de notas informativas y reportajes. Además, en esta estrategia también intervienen las redes sociales (Facebook, Twitter) del Instituto y del CdE, lo que le da una proyección internacional a la encuesta.

En conclusión, se puede afirmar que la encuesta de victimización en México es una buena práctica a nivel internacional, ya que responde a las necesidades de información tanto del público nacional como internacional. Adicionalmente, la ENVIPE se encuentra constantemente en revisión con el objetivo de cubrir las nuevas necesidades de información mediante el apoyo técnico de la Oficina de las Naciones Unidas contra la Droga y el Delito, a través del Centro de Excelencia para Información Estadística de Gobierno, Seguridad Pública, Victimización y Justicia (CdE).

Capítulo 2.

Modelo factorial para el análisis multivariado

2.1 Introducción

En el presente capítulo se expondrá el método científico de análisis de componentes principales. Entre sus fortalezas, destaca su capacidad de resumir un conjunto de indicadores básicos preservando la proporción máxima posible de la variación total de los datos. Indica que las mayores ponderaciones son asignadas a los indicadores básicos que muestran la mayor variación entre la población mayor de 18 años que ha sido víctima por delitos de robo o asalto en calle o transporte público, extorsión, fraude, robo total o parcial de vehículo, amenazas verbales, robo en casa habitación, lesiones, entre otros. En contrapartida, las debilidades del método son que las correlaciones no representan necesariamente la influencia real de los indicadores básicos en el fenómeno que está siendo medido y es sensible a modificaciones en los datos.

2.2 Planteamiento del análisis factorial

El análisis factorial es el nombre común que se le da a uno de los métodos estadísticos multivariantes que tiene como propósito principal analizar la estructura de las interrelaciones entre un gran número de variables sin hacer distinción entre variables dependientes e independientes. A partir de esta información se realiza un cálculo de dimensiones latentes que se conocen como factores buscando explicar las interrelaciones entre sí. (De la Garza, 2013)

Este análisis es una técnica de reducción de datos, el propósito es sintetizar las interrelaciones observadas entre un conjunto de variables en una forma concisa y segura como una ayuda a la construcción de nuevos conceptos y teorías. Para ello, se utiliza un conjunto de variables aleatorias que se construyen, llamadas factores comunes, de forma que todas las covarianzas o correlaciones son explicadas por dichos factores y cualquier porción de la varianza inexplicada por los factores comunes se asigna a términos de error residuales llamados factores únicos o específicos.

La importancia de este método consiste en que realiza tanto un análisis exploratorio como confirmatorio. El análisis exploratorio se caracteriza porque no se conocen *a priori* el número de factores y es en la aplicación empírica donde se determina este número. Por el contrario, en el análisis de tipo confirmatorio los factores están fijados a priori, utilizándose contrastes de hipótesis para su corroboración.

2.2.1 Componentes principales

En términos generales, los componentes principales analizan la estructura de las correlaciones entre un gran número de variables mediante la definición de un conjunto de dimensiones subyacentes comunes, conocidas como factores. Así, el analista puede identificar primero las dimensiones separadas por la estructura y después determinar el grado en que se explica cada variable por cada dimensión.

La formulación del análisis factorial (AF)⁴ es muy similar al análisis de componentes principales (CP), ambas son técnicas estadísticas del análisis multivariado, para identificar el menor número de factores que pueden ser usados para representar de forma resumida la información contenida en un conjunto numeroso de variables interrelacionadas, pero conceptualmente tienen diferentes significados.

La diferencia entre AF y CP es “si la matriz de correlación existe, las CP también existen, mientras que el modelo AF podría ser aceptado o no mediante un test estadístico”. Mediante la matriz de correlación en las variables, el AF se obtiene de los factores comunes.

Este método consiste en estimar las puntuaciones factoriales mediante las puntuaciones tipificadas de las primeras componentes principales y la matriz de cargas factoriales mediante las correlaciones de las variables originales con dichas componentes. A pesar de que siempre proporciona una solución, puede llevar a estimadores muy sesgados de la matriz de cargas factoriales, particularmente, si existen variables con comunalidades bajas.

Este modelo asume que variación específica es tan pequeña que la considera como cero y desprecia la variación aleatoria. Por lo tanto, la variación total se considera igual a la variación común. Esto significa que el máximo valor que puede tener la variación común es 1. En un análisis de tipo confirmatorio se supone que el factor está formado de una combinación lineal de las variables originales.

$$F_{ij} = U_1X_{1j} + U_2X_{2j} + \dots + U_kX_{kj}$$

Donde:

F = es el factor 1 de la observación j .

U = representa la importancia o peso relativo que cada variable estandarizada tiene con respecto al factor encontrado.

X = es la variable.

J = es el número de la observación.

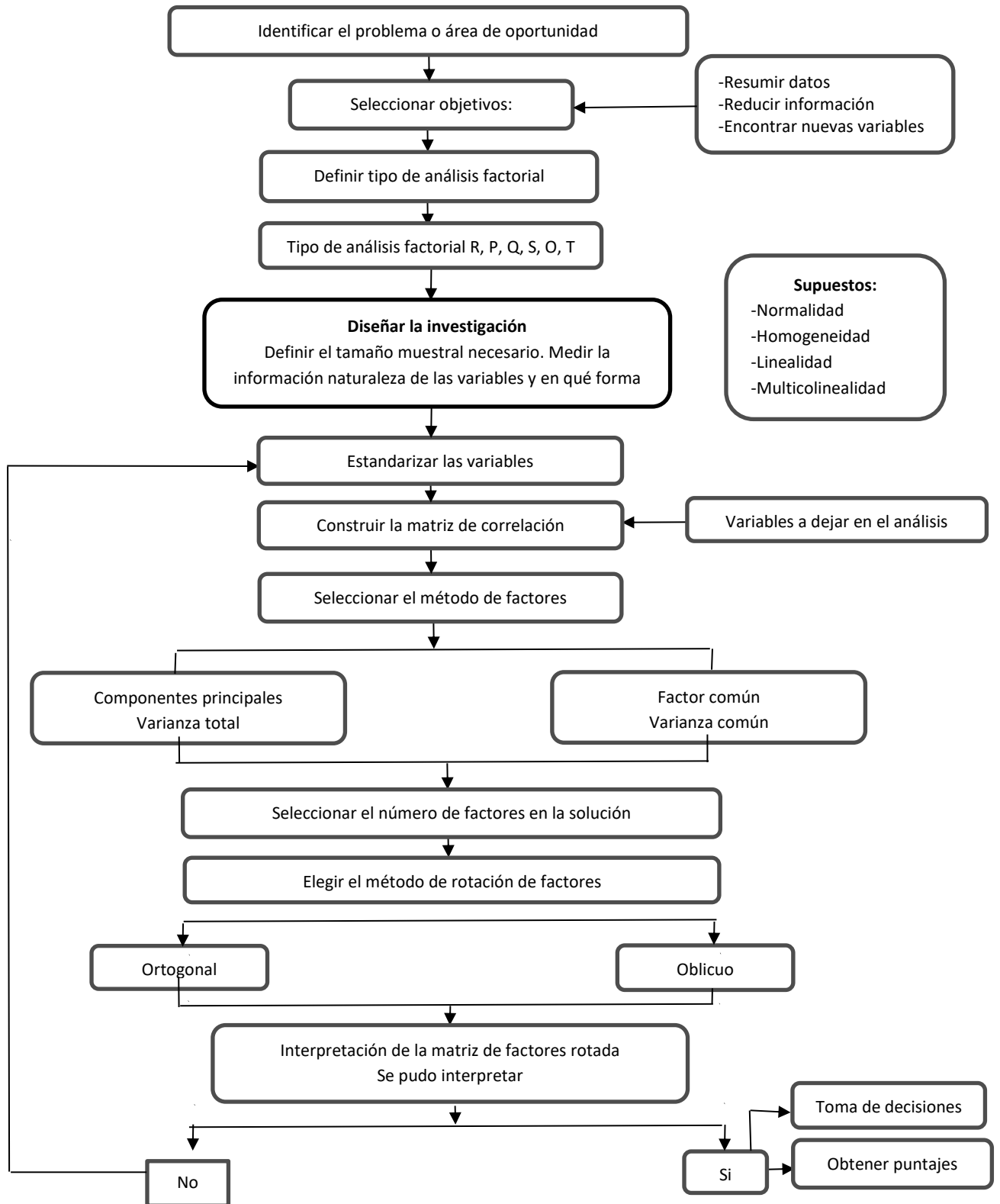
K = número de variables.

2.2.2 Diseño de la investigación

En éste debe abordarse la selección de las variables a analizar, así como la de los elementos de la población en la que dichas variables van a ser observadas. Y aunque se pueden realizar análisis factoriales con variables discretas y/o ordinales, lo habitual será que las variables sean cuantitativas continuas. Por lo que es importante, en todo caso, que dichas variables recojan los aspectos más esenciales de la temática sobre la victimización y percepción de la seguridad pública que se desea investigar y su selección deberá estar marcada por la teoría subyacente al problema. Aquellas variables que no vengan fundamentadas por los aspectos teóricos del problema no tiene caso incluirlas, debido a que se corre el riesgo de que los resultados obtenidos ofrezcan una estructura factorial difícil de entender y con escaso contenido teórico relevante.

4 Método multivariante que pretende expresar p variables observables como combinación lineal de m variables hipotéticas o latentes, conocidas como factores.

Figura 2.1
Pasos necesarios para realizar un análisis factorial



Fuente: Elaboración propia a partir de Jorge De la Garza García et aut. Análisis estadístico multivariante. Un enfoque teórico y práctico. Editorial Mc Graw Hill. México. 2013

2.3 Objetivos específicos

El objetivo principal es resumir la información asociada al nivel de victimización por delitos perpetrados en los hogares y en las personas para que su caracterización se pueda realizar más fácilmente, reduciendo las dimensiones o variables.

Los objetivos específicos serán:

- i. Identificar un conjunto de características que se encuentran latentes (que no se detectan fácilmente) dentro de un gran conjunto de variables.
- ii. Encontrar factores comunes que inciden en los delitos a mayores de 18 años, por entidad federativa, sexo de la víctima y tipología; así como aquellos que inciden en la percepción sobre seguridad pública.
- iii. Construir un índice que mida el nivel de victimización por delitos perpetrados en los hogares y en las personas, por entidad federativa y tipo de delito.

2.3.1 Reducir información

El análisis de componentes principales en resumen es una técnica estadística usada para identificar un número relativamente pequeño de factores que puede ser usado para representar en forma resumida la información contenida en un conjunto de variables interrelacionadas. Basado en este contexto, las variables relacionadas con los delitos de robo o asalto en calle o transporte público, extorsión, fraude, robo total o parcial de vehículo, amenazas verbales, robo en casa habitación, lesiones y otros delitos, pueden ser utilizadas para obtener un indicador del nivel de victimización por entidad federativa.

2.4 Etapas del análisis factorial

La formulación del análisis factorial es muy similar al análisis de componentes principales, ambas son técnicas estadísticas del análisis multivariado, para identificar el menor número de factores que pueden ser usados para representar de forma resumida la información contenida en un conjunto numeroso de variables interrelacionadas, pero conceptualmente tienen diferentes significados.⁵

Las diferencias entre el análisis factorial y componentes principales es si la matriz de correlación existe, las componentes principales también existen, mientras que el modelo del análisis factorial podría ser aceptado o no mediante un *test* estadístico. Mediante la matriz de correlación de las variables, el análisis factorial se obtiene de los factores comunes.

Dillon y Goldstein (citado por De la Garza, 2013) aseguran que el hecho de que componentes principales no hace ninguna distinción entre las fuentes de variación tiene otras implicaciones empíricas y conceptuales. Para ver estas consecuencias consideran la forma algebraica de las componentes principales, los factores W no observables se expresan como funciones de las variables observables:

⁵ Método multivariante que pretende expresar p variables observables como combinación lineal de m variables hipotéticas o latentes, conocidas como factores.

Dónde F_1, \dots, F_k ($k < p$) son los factores comunes y u_1, \dots, u_p los factores únicos o específicos y los coeficientes $\{a_{ij}; i = 1, \dots, p; j = 1, \dots, k\}$ las cargas factoriales. Los factores comunes están a su vez estandarizados ($E(F_i) = 0; Var(F_i) = 1$), los factores específicos tienen media 0 y están correlacionados ($E(u_i) = 0; Cov(u_i, u_j) = 0$ si $i \neq j; j, i = 1, \dots, p$) y ambos tipos de factores están correlacionados ($Cov(F_i, u_j) = 0; \forall i = 1, \dots, k; j = 1, \dots, p$)

Si, además, los factores están correlacionados ($Cov(F_i, F_j) = 0; si i \neq j; j, i = 1, \dots, k$) tenemos un modelo con factores ortogonales. En caso contrario el modelo se dice que es de factores oblicuos. Expresado en forma matricial

$$x = Af + u \Leftrightarrow X = FA' + U.$$

Donde, $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $f = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix}$, $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ X es la matriz de datos, $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{p1} & a_{p2} & \dots & a_{pk} \end{bmatrix}$ es la matriz de cargas factoriales y $F = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1k} \\ f_{21} & f_{22} & \dots & f_{2k} \\ \dots & \dots & \dots & \dots \\ f_{p1} & f_{p2} & \dots & f_{pk} \end{bmatrix}$ es la matriz de puntuaciones factoriales.

Utilizando las hipótesis anteriores se tiene que $Var(X_i) = \sum_{j=1}^k a_{ij}^2 + \psi_i = h_i^2 + \psi_i; i = 1, \dots, p$: donde $h_i^2 = Var(\sum_{j=1}^k a_{ij} F_j)$ y $\psi_i = Var(u_i)$ reciben los nombres de comunalidad y especificidad de la variable X_i , respectivamente.

Por lo tanto, la varianza de cada una de las variables analizadas puede descomponerse en dos partes: una, la comunalidad h_i^2 que representa la varianza explicada por los factores comunes y otra la especificidad ψ_i que representa la parte de la varianza específica de cada variable.

Además, se tiene que $Cov(X_i, X_j) = Cov(\sum_{j=1}^k a_{ij} F_j, \sum_{j=1}^k a_{lj} F_j) = \sum_{j=1}^k a_{ij} a_{lj} \forall i \neq l$ por lo que son los factores comunes los que explican las relaciones existentes entre las variables del problema. Es por esta razón que los factores que tienen interés y son susceptibles de interpretación experimental son los factores comunes. Los factores únicos se incluyen en el modelo de acuerdo a la imposibilidad de expresar, en general, p variables en función de un número más reducido k de factores.

2.4.1.1 Estandarización de variables

Es importante contar con una idea más o menos clara de cuáles son los factores comunes para medir y elegir las variables de acuerdo a los mismos. La muestra debe ser representativa de la población objeto de estudio y del mayor tamaño posible. Los resultados del análisis no tienen por qué ser invariantes a cambios de origen y escala por lo que si las unidades de medida de las variables no son comparables se deben estandarizar los datos antes de realizar el análisis.

Después de obtener la matriz de componentes, se definen las puntuaciones factoriales mediante una estimación para cada uno de los sujetos en cada factor extraído, con el objeto de valorar la situación que tiene ese sujeto frente a las variables construidas (Factores).

El cálculo de las puntuaciones factoriales son las puntuaciones que tienen los componentes principales para cada uno de los sujetos de la muestra, dichas puntuaciones se calculan mediante la expresión:

$$X_{ij} = a_{i1} \cdot Z_{1j} + \dots + a_{ip} \cdot Z_{pj} = \sum_{s=1}^k a_{is} \cdot Z_{sk}.$$

Donde:

a_{is} – son los coeficientes.

Z_{sk} – son los valores estandarizados.

Al utilizar el método de extracción de factores, las puntuaciones obtenidas se calculan a partir de la solución factorial, esta solución es ortogonal⁶ por lo que las puntuaciones también lo son.

2.4.1.2 Matriz de correlación reproducida

Para calcular las correlaciones es necesario obtener las varianzas y covarianzas de las variables y después encontrar las correlaciones, pero si las variables se encuentran estandarizadas no es necesario este paso ya que las varianzas y covarianzas de variables estandarizadas son las correlaciones; esta es una de las ventajas que se tiene cuando se estandarizan las variables [De la Garza].

La matriz de correlación entre las variables observadas es usualmente un arreglo de forma cuadrática:

$$R_p = \begin{pmatrix} 1 & r_{12} \dots r_{1p} \\ r_{21} & 1 \dots r_{2p} \\ \dots & \dots \dots \\ r_{p1} & r_{p2} \dots 1 \end{pmatrix}.$$

Donde $r_{ij} = \frac{S_{ij}}{(S_i S_j)^{1/2}}$

S_i , $i = (1, \dots, p)$ son las varianzas de cada una de las variables observadas

S_{ij} , $i = (1, \dots, p)$ y $j = (1, \dots, p)$ son las covarianzas de par en par de las variables observadas.

Como se observa, los elementos de la diagonal presentan una correlación igual a 1, mientras que el resto de los elementos sus correlaciones son menores o iguales a 1. El valor numérico de las correlaciones oscila entre [+1,-1], cuando la correlación relativamente cercana a +1 o -1 significa que existe una correlación entre ambas variables, mientras que, si da cero, significa que ambas variables no tienen correlación alguna.

Adicionalmente, otra característica que presenta una matriz de correlación es su simetría, es decir, si se divide la matriz por la diagonal se observa que los elementos del triángulo inferior son iguales a los elementos del triángulo superior.

La matriz de correlación indica cuales variables posiblemente quedaran agrupadas en el mismo factor y cuales posiblemente no lo hagan. La matriz de correlación sirve para cuando las variables se identifican con más de un factor o característica y se tiene que deslindar para que cada variable se relacione con un solo factor.

⁶ Independencia matemática de ejes factoriales entre sí (i.e., ángulos rectos o ángulos de 90 grados).

2.4.2 Análisis factorial confirmatorio

La técnica de extracción de factores puede utilizarse en un análisis confirmatorio o exploratorio. El modelo de componentes principales es útil si el objetivo es confirmar una teoría o hipótesis previamente establecida, lo que se llamará confirmatorio y el modelo de factor común se enfoca en un análisis de tipo exploratorio, pues se desea estudiar dentro del conjunto de datos esa estructura latente que hace que las variables muestren interrelación.

Las variables extraídas se deberán estandarizar y la varianza de cada una de ellas será de 1. La extracción de factores busca aspectos comunes entre variables para agruparlas y la dispersión o variación de los datos indica posibles similitudes entre las variables.

La matriz de cargas factoriales juega un papel destacado a la hora de interpretar el significado de los factores y, si éstos son ortogonales, cuantifican el grado y tipo de la relación entre éstos y las variables originales. Para resolver este problema están los procedimientos de *Rotación de Factores* que buscan obtener, a partir de la solución inicial, unos factores cuya matriz de cargas factoriales los hagan más fácilmente interpretables.

Dichos métodos intentan aproximar la solución obtenida al Principio de Estructura Simple (Thurstone, 1935) según la cual la matriz de cargas factoriales debe reunir las siguientes características:

- 1) cada factor debe tener unos pocos pesos altos y los otros próximos a cero;
- 2) cada variable no debe estar saturada más que en un factor;
- 3) no deben existir factores con la misma distribución, es decir, dos factores distintos deben presentar distribuciones diferentes de cargas altas y bajas.

De esta forma y dado que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de las variables. Examinando las características de las variables de un grupo asociado a un determinado factor se pueden encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a esos rasgos comunes.

Si se consiguen identificar claramente estos rasgos, se habrá dado un paso importante, ya que con los factores comunes no sólo se reducirá la dimensión del problema, sino que también se conseguirá desvelar la naturaleza de las interrelaciones existentes entre las variables originales.

Existen dos formas básicas de realizar la rotación de factores: la Rotación Ortogonal y la Rotación Oblicua, según que los factores rotados sigan siendo ortogonales o no. Conviene advertir que tanto en la rotación ortogonal, como en la rotación oblicua la comunalidad de cada variable no se modifica, es decir, la rotación no afecta a la bondad de ajuste de la solución factorial: aunque cambie la matriz factorial, las especificidades no cambian y por tanto, las comunalidades permanecen inalteradas. Sin embargo, cambia la varianza explicada por cada factor, luego los nuevos factores rotados no están ordenados de acuerdo con la información que contienen cuantificada a través de su varianza.

2.4.2.1 Hipótesis y supuestos.

Las hipótesis propuestas son:

Hipótesis 1. La prueba KMO (Keyser, Meyer, Olkin) debe ser mayor a 0.75 para construir un índice de consumo por medio del análisis de factores.

Hipótesis 2. La medida de adecuación muestral para cada una de las variables debe ser mayor a 0.75, es decir, cada una de las variables deberá ser compatible con el modelo.

Hipótesis 3. Se rechaza la hipótesis nula en el Test Bartlett, es decir, que la matriz de correlaciones es significativamente diferente de una matriz idéntica y las variables están incorrelacionadas.

Hipótesis 4. La primera componente explica el modelo muy cercano al 93.91% de la varianza contenida en las variables originales.

Hipótesis 5. Se puede expresar mediante la combinación lineal de una ecuación basada en los coeficientes de los factores.

2.4.2.2 Test de esfericidad de Bartlett

Es una manera particular de analizar la matriz de correlación. El test de esfericidad de Bartlett se obtendrá a partir de una transformación del determinante de la matriz de correlación, es decir, si el determinante de la matriz de correlación es igual a uno, quiere decir que las intercorrelaciones entre las variables son cero, lo que significa que las variables no están intercorrelacionadas por lo cual el Análisis Factorial no sería el método apropiado de utilizar.

La prueba de esfericidad de Bartlett permite contrastar que en la matriz de correlación exista una relación entre las variables, es decir, que exista cierto grado de multicolinealidad, ya que la técnica de análisis de factores identifica las variables que están interrelacionadas. En caso contrario, la matriz de correlación sería una matriz identidad y, por lo tanto, no tendría sentido un análisis de factores.

Sea R_ρ la matriz de correlación de las variables observadas se prueba que las correlaciones no son nulas y, por tanto, hay expectativas de tener éxito en la extracción de los factores. Si se confirma la hipótesis nula significa que las variables no están intercorrelacionadas; por lo tanto, se plantea:

$$H_0: |R_\rho| = 1 \text{ o } R_\rho = I \text{ vs. } H_a: |R_\rho| \neq 1.$$

El estadístico de dicho Test está dado por⁷:

$$d_{R_\rho} = - \left[n - 1 - \frac{1}{6}(2p + 5) \right] \log |R_\rho| = - \left[n - \frac{2p+11}{6} \right] \sum_{j=1}^p \log(\lambda_j).$$

Donde:

n – número de individuos de la muestra
j=1,..., p que son los valores propios R_ρ

Bajo la hipótesis nula, el estadístico tiende a ser una distribución ji-cuadrada con $P(P-1)/2$ grados de libertad, es decir, si una matriz de correlación es igual a la idéntica entonces significa que las intercorrelaciones entre las variables son igual a cero. Si con el test Bartlett se obtiene valores altos de χ^2 , esto significa que hay variables con correlaciones altas, por lo tanto, la hipótesis nula se rechaza. En caso de no rechazar la hipótesis nula significa que las variables no están intercorrelacionadas y en este supuesto debería de reconsiderarse la aplicación de un análisis de componentes.

⁷ Harman, H. H. (1980). Análisis factorial moderno. Madrid: Saltés.531 páginas.

2.4.2.3 Medida de la adecuación muestral

Esta medida también sirve para indicar si se pudiera utilizar el análisis de factores, pero también permite conocer qué tan fuerte y adecuada sería la posible solución a encontrar, es decir, entre más grande sea el valor obtenido tras utilizar esta medida más fuerte es la solución, por lo que se definen rangos que determinan el valor de mi evaluación.

Cuadro 2.1
Evaluación para la medida de adecuación

KMO	Evaluación
De .90 en adelante	Excelente
De .80 a .90	Bueno
De .70 a .80	Aceptable
De .60 a .70	Regular
De .50 a .60	Bajo
Menor a .50	Inaceptable

Fuente: De la Garza García Jorge et al. Análisis estadístico multivariante. Un enfoque teórico y práctico. Editorial Mc Graw Hill. México. 2013

La medida de adecuación muestral es una medida global que indica si se llevará a cabo el análisis de factores, qué tan fuerte y adecuada sería la posible solución a encontrar, mientras más grande es este valor, la solución es más fuerte, lo óptimo es que $MASg \geq 0.5$.

Este coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto. Si las variables comparten factores comunes, el coeficiente de correlación parcial debe ser bajo entre pares de variables, puesto que se eliminan los efectos lineales de las otras variables.

Las correlaciones parciales son estimaciones de las correlaciones entre los factores únicos y deberían ser próximos a cero cuando el análisis de componentes principales es adecuado, ya que, estos factores se supone que están incorrelacionados entre sí. Por lo tanto, si existe un número elevado de coeficientes de este tipo distintos de cero es señal de que las hipótesis del modelo factorial no son compatibles con los datos.

2.4.2.4 Regla de KMO

Esta regla permite determinar aquellos factores a conservar, mediante el cálculo de valores propios de la matriz de correlaciones y tomar como número de factores el número de valores propios superiores a la unidad, comúnmente esta regla permite establecer un límite inferior. Para obtener un límite superior habría que aplicar la misma regla estableciendo como límite 0.75.

La regla consiste en calcular los valores propios de la matriz de correlaciones R y tomar como número de factores el número de valores propios superiores a la unidad. Este criterio es una reminiscencia del análisis de componentes principales y se ha comprobado en simulaciones que, generalmente, tiende a

infraestimar el número de factores por lo que se recomienda su uso para establecer un límite inferior. Un límite superior se calcularía aplicando este mismo criterio, pero tomando como límite 0.75.

La medida de adecuación de la muestra KMO determina que el coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto. Las correlaciones parciales entre los factores únicos deben ser próximos a cero.

La medida de adecuación de la muestra KMO propuesta por Kaiser, Meyer y Olkin, la cual está dada por:

$$KMO = \frac{\sum_{j \neq i} \sum_{i \neq j} r_{ij}^2}{\sum_{j \neq i} \sum_{i \neq j} r_{ij}^2 + \sum_{j \neq i} \sum_{i \neq j} r_{ij(p)}^2}.$$

Donde $r_{ij(p)}$ es el coeficiente de correlación parcial entre las variables X_i, X_j eliminando la influencia del resto de las variables.

KMO puede tomar valores entre [0,1], se utiliza para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial de forma que, cuanto más pequeño sea el valor, mayor es el valor de los coeficientes de correlación parcial $r_{ij(p)}$ y por lo tanto menos deseable es realizar un análisis de componentes principales.

El índice se puede interpretar como la potencialidad del análisis componentes principales para extraer la variabilidad de las variables originales. Por esta razón, Kaiser, Mayer y Olkin, aconsejan los siguientes criterios para establecer la idea de realizar el análisis de componentes principales:

$KMO \geq 0.75$ – Es excelente.

$0.75 > KMO \geq 0.5$ – Es aceptable.

$KMO < 0.5$ – Es inaceptable.

Dado lo anterior, la medida KMO calculada para la generación del índice de percepción de violencia de género deberá arrojar como resultado un valor superior a 0.5 y se aprobará la H_1 .

2.4.2.5 Criterio de porcentaje de varianza

Se deriva del análisis de componentes principales y consiste en tomar como número de factores el número mínimo necesario para que el porcentaje acumulado de la varianza explicado alcance un nivel satisfactorio que suele ser del 75% o el 80%. Carece de justificación teórica o práctica cuando se aplica este criterio a la matriz de varianzas o covarianzas.

También es una reminiscencia del análisis de componentes principales y consiste en tomar como número de factores el número mínimo necesario para que el porcentaje acumulado de la varianza explicado alcance un nivel satisfactorio que suele ser del 75% o el 80%. Tiene la ventaja de poderse aplicar también cuando la matriz analizada es la de varianzas y covarianzas pero no tiene ninguna justificación teórica ni práctica.

El criterio más utilizado para escoger el número de factores a tomar es con base en el porcentaje de varianza total explicada por cada factor, y cuando este llega a un porcentaje acumulado considerablemente alto, normalmente cerca del 80%, significa que el número de factores es suficiente.

2.4.2.6 Extracción de factores

Es la selección reducida de factores que puedan representar las variables originales, posterior a que se haya determinado que el Análisis Factorial es la técnica adecuada para analizar los datos. Cuando se intenta cuantificar la matriz de cargas factoriales que es la que explica la matriz de datos en función de los factores comunes, se logra obtener la matriz correlacional poblacional de las variables (identidad fundamental del análisis factorial) que está dada por la matriz diagonal de las especificidades. Existen diversos métodos para la extracción de factores, cada uno con ventajas e inconvenientes.

- A. Problema de grados de libertad. Surge cuando el proceso de estimación requiere que el número de ecuaciones sea mayor o igual que el número de parámetros a estimar.
- B. Problema de no unicidad de la solución. En caso de que no requiera un número mayor o igual de ecuaciones que el número de parámetros a estimar, se debe considerar que la solución a la matriz de cargas factoriales no es única (cualquier transformación ortogonal de la matriz de cargas factoriales es una solución). Por lo tanto, habrá que cobrar consciencia que salvo las rotaciones ortogonales el modelo es único.

2.4.2.7 Determinación del número de factores

Cuando se tienen factores que contribuyen poco, muy poco o nada a la solución, se deberá determinar aquellos factores a conservar para intentar cumplir con el *principio de parsimonia*.

La aplicación de cada uno de los siguientes criterios deberá hacerse en la información reportada por la matriz no rotada, y como guía para determinar una primera solución al problema. La solución definitiva no necesariamente será la obtenida por el grado de los criterios mencionados; estas se conocerán al analizar soluciones con uno o dos factores más o menos al número de factores obtenidos en la solución inicial. Aquella solución que agrupe lógicamente a las variables será la definitiva.

El objetivo para determinar el número de factores como solución inicial son los siguientes:

- Criterio a priori. Se utiliza en aquellos casos en donde el investigador desea probar alguna teoría o hipótesis y de antemano conoce cuantos factores se deben tener en la solución; el investigador decide el número de factores a pedir con base a su hipótesis. Se recomienda si se usa la técnica de análisis de factores para un análisis confirmatorio.
- Criterio de raíz latente. En este se pueden tomar en cuenta dos consideraciones: En caso de que los datos no estén estandarizados la idea es retener a los factores cuya raíz característica exceda la media de las raíces característica. La raíz característica es la variación explicada por cada factor. Cuando los datos están estandarizados se considera que un factor debe ser retenido en la solución si su raíz característica es mayor a 1. Debido a que el eigenvalor o raíces características representan las varianzas y las varianzas de las variables estandarizadas es igual a 1, un factor con un valor eigenvalor menor a 1, no es importante o se considera que no tiene la cantidad de información significativa captada; lo lógico es mantener a los factores que tengan mayor información que cualquiera de las originales. Se recomienda este criterio si el objetivo del análisis de factores es para un análisis exploratorio.

- Criterio del porcentaje de variación explicada acumulada. Mediante este criterio se considera que n factores deben manejarse como solución inicial, si el porcentaje de variación explicada acumulada se encuentra en un rango entre 60 y 95%. Es decir, a través de este criterio se está dispuesto a perder cuando mucho el 40% y cuando menos 5% de la información.
- Criterio Scree Test. La idea es determinar que, si los factores son importantes, tendrán una varianza grande. Al extraer los factores, el primero contabiliza la mayor variación, el segundo que se extrae tiene menor variación que el primero y así sucesivamente. En virtud de esta manera natural en que se extraen, para la aplicación de este criterio se realiza una gráfica en la que se ubica el número del factor en el eje de las X y los valores de la varianza con los eningenvalores en el eje de las Y.

2.4.2.8 Cálculo Puntuaciones factoriales.

Una vez que se determinaron los factores rotados, de deberá calcular la matriz de puntuaciones factoriales. Se llama puntaje a los datos con los que se manejaría cada factor como variable. Los puntajes se pueden obtener a través de tres enfoques

1. Preciso
 2. Aproximado
 3. Medio
1. Preciso. En este enfoque, existen tres criterios para obtener los coeficientes estandarizados que forman ecuaciones lineales para la extracción de los puntajes y son: regresión, Anderson-Rubin y Bartlett.

En los tres criterios, se involucran todas las variables y consiste en obtener los puntajes respectivos al sustituir los datos estandarizados de la matriz de coeficientes estandarizados en la ecuación que define al factor. Dicha ecuación está formada por una combinación lineal de las variables que se manejan en la investigación; entonces, los datos obtenidos se sustituyen en la ecuación y resulta un puntaje del factor, es decir, habrá tantos puntajes como datos se tengan.

La diferencia entre cada criterio es: en el de Anderson-Rubin los puntajes que se alcanzan son de variables sin relación (ortogonales) y estandarizados. En el de regresión los puntajes de factores pueden estar correlacionados y en el de Bartlett se aplica el método de máxima verosimilitud, bajo el supuesto de que los factores deben presentar normalidad. También se le conoce como método largo.

2. Aproximado. Consiste en utilizar como puntajes los datos originales de la variable que se haya identificado más con el factor, para esto es necesario volver a la matriz rotada y encontrar la variable que tenga la mayor carga pero por factor. También se le conoce como método corto.
3. Promedio. Consiste en utilizar como puntajes los promedios de las variables que se agruparon con cada factor. Para calcular los puntajes del factor 1 es necesario volver a la matriz rotada y saber cuáles son las variables que se agruparon con dicho factor. Con esta información se vuelve a los datos originales (si no están disponibles o si no se tienen las mismas unidades, se

pueden utilizar los valores de las variables estandarizados) y se encuentra el promedio de dichas variables.

Después de obtener la matriz de componentes, se definen las puntuaciones factoriales mediante una estimación para cada uno de los sujetos en cada factor extraído, con el objeto de valorar la situación que tiene ese sujeto frente a las variables construidas (factores).

El cálculo de las puntuaciones factoriales son las puntuaciones que tienen los componentes principales para cada uno de los sujetos de la muestra, dichas puntuaciones se calculan mediante la expresión:

$$X_{ij} = a_{i1} \cdot Z_{1j} + \dots + a_{ip} \cdot Z_{pj} = \sum_{s=1}^k a_{is} \cdot Z_{sk}.$$

Dónde:

a_{is} – son los coeficientes

Z_{sk} – son los valores estandarizados

Al utilizar el método de extracción de factores, las puntuaciones obtenidas se calculan a partir de la solución factorial, esta solución es ortogonal⁸ por lo que las puntuaciones también lo son.

2.5 Validación del modelo

Este es el último paso en el análisis factorial, que implica estudiar la validez del modelo mediante el análisis de la bondad de ajuste del modelo y la generalización de sus resultados. Afifi, A., May S. & Clark V. en su libro *Practical multivariate analysis*, describen el análisis de componentes principales de forma general:

Supongamos que tenemos N observaciones de X_1, X_2 donde:

$$x_1 = X_1 - \bar{X}_1, \quad x_2 = X_2 - \bar{X}_2.$$

Se desea crear dos nuevas variables C_1, C_2 llamadas componentes principales. Estas nuevas variables son función lineal de x_1, x_2 por lo tanto se pueden escribir como:

$$C_1 = a_{11}x_1 + a_{12}x_2.$$

$$C_2 = a_{21}x_1 + a_{22}x_2.$$

Las medias y las varianzas de las N observaciones para C_1, C_2 son:

$$\text{mean } C_1 = \text{mean } C_2 = 0.$$

$$\text{Var } C_1 = a_{11}^2 S_1^2 + a_{12}^2 S_2^2 + 2a_{11}a_{12}rS_1S_2.$$

$$\text{Var } C_2 = a_{21}^2 S_1^2 + a_{22}^2 S_2^2 + 2a_{21}a_{22}rS_1S_2.$$

Donde:

$$S_i^2 = \text{Var}X_i.$$

⁸ Independencia matemática de ejes factoriales entre sí (i.e., ángulos rectos o ángulos de 90 grados).

Los a_{ij}^2 son los coeficientes⁹ se escogen satisfactoriamente por tres requisitos:

1. $Var C_1$ es el más largo posible.
2. Las N valores de C_1 y C_2 son incorrelacionados.
3. $a_{11}^2 + a_{12}^2 = a_{21}^2 + a_{22}^2 = 1$.

Este modelo puede ser aplicado para la generación de un índice de victimización; sin embargo, es necesario encontrar las variables que puedan ser incluidos en el modelo. [Afifif, May & Clark, 2012]

2.6 Conclusión del capítulo

En este capítulo se estudiaron las técnicas para un análisis confirmatorio o exploratorio mediante el modelo de componentes principales cuyo objetivo es confirmar una teoría o hipótesis previamente establecidas, por lo que se llamará confirmatorio. El modelo de factor común se enfoca a un análisis de tipo exploratorio, pues se desea estudiar dentro del conjunto de datos esa estructura latente que hace que las variables muestren interrelación.

Las variables extraídas deberán estandarizar y la varianza de cada una de ellas es 1. La extracción de factores busca aspectos comunes entre variables para agruparlas y la dispersión o variación de los datos indica posibles similitudes entre las variables.

La matriz de cargas factoriales juega un papel destacado a la hora de interpretar el significado de los factores y, si éstos son ortogonales, cuantifican el grado y tipo de la relación entre éstos y las variables originales. Para resolver este problema están los procedimientos de Rotación de Factores que buscan obtener, a partir de la solución inicial, unos factores cuya matriz de cargas factoriales los haga más fácilmente interpretables.

Dichos métodos intentan aproximar la solución obtenida según el cual la matriz de cargas factoriales debe reunir las siguientes características: I. Cada factor debe tener unos pocos pesos altos y los otros próximos a cero; II. Cada variable no debe estar saturada más que en un factor; y III. No deben existir factores con la misma distribución, es decir, dos factores distintos deben presentar distribuciones diferentes de cargas altas y bajas.

De esta forma y dado que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de las variables. Examinando las características de las variables de un grupo asociado a un determinado factor se pueden encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a esos rasgos comunes.

Si se consiguen identificar claramente estos rasgos, se habrá dado un paso importante, ya que con los factores comunes no sólo se reducirá la dimensión del problema, sino que también se conseguirá desvelar la naturaleza de las interrelaciones existentes entre las variables originales.

⁹ La solución matemática para los coeficientes fue originalmente derivada por Hottelling (1993).

Capítulo 3.

Aplicación del modelo factorial a la ENVIPE 2020

3.1 Introducción

El análisis factorial, supone que cada una de las p variables observadas, es función de un número m factores comunes ($m < p$) más un factor específico o único. Tanto los factores comunes como los específicos no son observables y su determinación e interpretación es el resultado de análisis factorial. Estas variables hipotéticas suponen: i) que los factores comunes son variables con medida cero y varianza uno. Además, se suponen incorrelacionados entre sí; ii) los factores son únicos son variables con medida cero, y sus varianzas pueden ser distintas; se supone que están incorrelacionados entre sí, de lo contrario la información obtenida de ellos estaría en los factores comunes; y, iii) los factores comunes y los factores únicos están incorrelacionados entre sí; esta hipótesis permite realizar inferencias que permitan distinguir entre los factores comunes y los específicos.

En este capítulo se trabajará la aplicación de la metodología de análisis factorial como técnica estadística de reducción de datos que permitirá explicar la variabilidad entre variables observadas en términos de un número menor de variables no observadas llamadas factores para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Estos grupos homogéneos se forman con las variables que correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros.

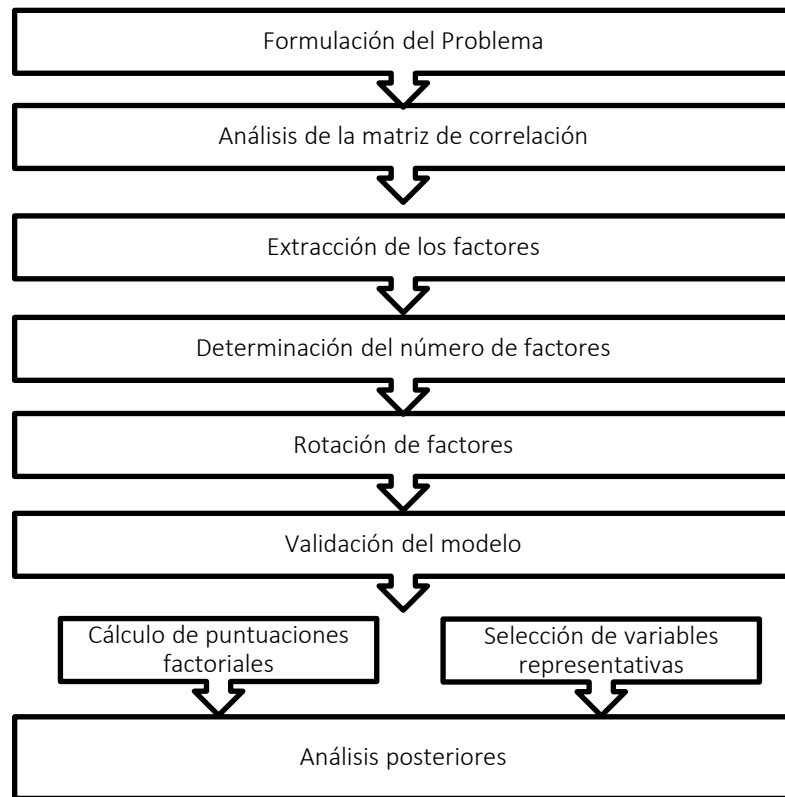
Las variables observadas se modelan como combinaciones lineales de factores más expresiones de error. Para obtener el *índice de victimización y percepción sobre seguridad pública* primeramente se obtendrá la matriz de correlaciones para ubicar las correlaciones entre todas las variables consideradas y posteriormente se aplicarán diversas pruebas estadísticas como el *test* de esfericidad de Bartlett y la prueba de Kaiser-Meyer-Olkin para medir la adecuación de la muestra.

3.2 Diseño de la investigación

El análisis factorial es el nombre común que se le da a uno de los métodos estadísticos multivariantes que tiene como propósito principal analizar la estructura de las interrelaciones entre un gran número de variables, sin hacer distinción entre variables dependientes e independientes. A partir de esta información se realiza un cálculo de dimensiones latentes que se conocen como factores buscando explicar la interrelación entre sí.

En este debe abordarse la selección de las variables a analizar, así como la de los elementos de la población en la que dichas variables van a ser observadas. Y aunque se pueden realizar análisis factoriales con variables discretas y/u ordinales, lo habitual será que las variables sean cuantitativas continuas. Por lo que es importante, en todo caso, que dichas variables recojan los aspectos más esenciales de la temática sobre el nivel el nivel de violencia y percepción de la seguridad pública que se desea investigar y su selección deberá estar marcada por la teoría subyacente al problema. Aquellas variables que no vengan fundamentadas por los aspectos teóricos del problema no tiene caso incluirlas, debido a que se corre el riesgo de que los resultados obtenidos ofrezcan una estructura factorial difícil de entender y con escaso contenido teórico relevante.

Figura 3.1
Pasos necesarios para realizar un análisis factorial



Fuente: Elaboración propia a partir de Jorge De la Garza García et aut. Análisis estadístico multivariante. Un enfoque teórico y práctico. Editorial Mc Graw Hill. México. 2013

3.2.1 Reducir información

El análisis de componentes principales en resumen es una técnica estadística usada para identificar un número relativamente pequeño de factores que puede ser usado para representar en forma resumida la información contenida en un conjunto de variables interrelacionadas. Basado en este contexto, las variables relacionadas con el nivel de victimización tales como:

- F1= Tasa de prevalencia delictiva por entidad federativa por cada cien mil habitantes, según sexo de la víctima 2020
- F2= Tasa de prevalencia delictiva por entidad federativa por cada cien mil habitantes, según tipo de delito 2020
- F3= Tasa de incidencia delictiva por entidad federativa de ocurrencia por cada cien mil habitantes y los cinco delitos más frecuentes 2020
- F4= Víctimas por entidad federativa, según lugar de victimización 2020
- F5= Población de 18 años y más por entidad federativa, según condición de victimización 2020

- F6= Población de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021
- F7= Población de hombres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021
- F8= Población de mujeres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021

3.2.2 Objetivo específico

Al utilizar el método de extracción de factores, las puntuaciones obtenidas se calculan a partir de la solución factorial, esta solución es *ortogonal*¹⁰ por lo que las puntuaciones también lo son. Mediante la matriz de coeficientes para el cálculo de las puntuaciones factoriales es posible construir una ecuación lineal de la primera componente para construir el índice de victimización y percepción de la seguridad pública cada una de las observaciones.

$$Iv = a_1Z_{I_1} + a_2Z_{I_2} + a_3Z_{I_3} + \dots + a_{n-1}Z_{I_{n-1}} + a_nZ_{I_n}$$

Sustituyendo Z_{I_i} que son los valores estandarizados, se deberá obtener el siguiente polinomio ortogonal:

$$Iv = a_1 \frac{(x_{I_1} - \bar{X}_{V_1})}{\sigma_{V_1}} + a_2 \frac{(x_{I_2} - \bar{X}_{V_2})}{\sigma_{V_2}} + a_3 \frac{(x_{I_3} - \bar{X}_{V_3})}{\sigma_{V_3}} + \dots + a_{n-1} \frac{(x_{I_{n-1}} - \bar{X}_{V_{n-1}})}{\sigma_{V_{n-1}}} + a_n \frac{(x_{I_n} - \bar{X}_{V_n})}{\sigma_{V_n}}$$

3.2.3 Integración de fuentes de datos

La ENVIPE es temporal ya que se consideró el 2019 para victimización y 2020 para percepción de la seguridad pública pues el periodo de levantamiento de información, debido a la emergencia sanitaria generada por el virus SARS-CoV2 (COVID-19), se realizó del 17 al 31 de marzo de 2019 y del 27 de julio al 04 de septiembre de 2020.

3.2.4 Herramienta de explotación

Para analizar los datos se utiliza el paquete estadístico *Statistical Product and Service Solutions SPSS*[®] por sus siglas en inglés). Este programa informático es muy utilizado en las ciencias sociales y las empresas de investigación de mercado para cálculos matemáticos porque se pueden trabajar fácilmente con bases de datos de gran tamaño; así como la recodificación de las variables y registros, según las necesidades del usuario.

El programa consiste en un módulo base y módulos anexos que se han actualizado constantemente con nuevos procedimientos estadísticos.

¹⁰ Independencia matemática de ejes factoriales entre sí (i.e., ángulos rectos o ángulos de 90 grados).

Las ventajas del programa son:

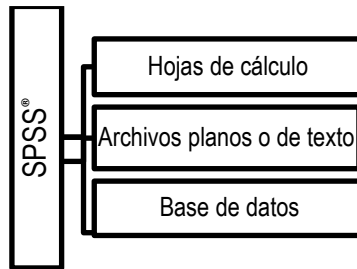
- Permite un importante ahorro de tiempo y esfuerzo, realizando en segundos un trabajo que requeriría horas e incluso días.
- Hace posibles cálculos más exactos, evitando los redondeos y aproximaciones del cálculo manual.
- Permite trabajar con grandes cantidades de datos, utilizando muestras mayores e incluyendo más variables.
- Permite trasladar la atención desde las tareas mecánicas de cálculo a las tareas conceptuales: decisiones sobre el proceso, interpretación de resultados, análisis crítico.

Por su parte, las desventajas son:

- El aprendizaje del manejo de paquetes de programas estadísticos requiere un cierto esfuerzo.
- A veces, la capacidad de cálculo del evaluador supera la capacidad para comprender el análisis realizado e interpretar los resultados.
- Lleva a veces a una sofisticación innecesaria, al permitir el empleo de técnicas complejas para responder a cuestiones simples.

SPSS® permite importar información proveniente de múltiples fuentes de datos como Excel®, Access® o archivos de texto. Para crear un archivo de datos, una opción es ingresar la información desde otro programa de computadora. Para la extracción de la información contenida en estos formatos, el programa tiene establecida una estructura de captura, basada en tres grupos.

Figura 3.2
Fuentes de datos para exportar en SPSS®



Fuente: Elaboración propia

Estos grupos son determinados por las características organizacionales de los datos dentro del archivo, cuyas propiedades son definidas por los programas que los generan. SPSS® permite extraer información de archivos de hojas de cálculo como Excel®.

Para el análisis de correlación factorial que se desarrollará en los apartados subsecuentes, se importaron los datos desde un archivo de Excel® y a las etiquetas se les asignó una medida, que en este caso fue la de «Escala». Posteriormente se comenzará el análisis factorial con la opción «reducción de dimensiones» y se seleccionaran los descriptivos de la matriz de correlaciones: coeficientes, niveles de significación, determinante, KMO y prueba de esfericidad de Bartlett, inverso, reproducida, anti-imagen. Dentro de rotación se elegirá el Método «Varimax» y en puntuaciones se seleccionará el Método Bartlett.

3.3 Elección del método estadístico

Las variables extraídas deberán estandarizar y la varianza de cada una de ellas es 1. La extracción de factores busca aspectos comunes entre variables para agruparlas y la dispersión o variación de los datos indica posibles similitudes entre las variables.

La matriz de cargas factoriales juega un papel destacado a la hora de interpretar el significado de los factores y, si estos son ortogonales, cuantifican el grado y tipo de la relación entre estos y las variables originales. Para resolver este problema están los procedimientos de *Rotación de Factores* que buscan obtener, a partir de la solución inicial, unos factores cuya matriz de cargas factoriales los haga más fácilmente interpretables.

Dichos métodos intentan aproximar la solución obtenida al Principio de Estructura Simple (Thurstone citado por De la Garza 2013) según el cual la matriz de cargas factoriales debe reunir las siguientes características:

1. Cada factor debe tener unos pocos pesos altos y los otros próximos a cero;
2. Cada variable no debe estar saturada más que en un factor;
3. No deben existir factores con la misma distribución, es decir, dos factores distintos deben presentar distribuciones diferentes de cargas altas y bajas.

De esta forma y dado que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de las variables. Examinando las características de las variables de un grupo asociado a un determinado factor se pueden encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a esos rasgos comunes.

Si se consiguen identificar claramente estos rasgos, se habrá dado un paso importante, ya que con los factores comunes no sólo se reducirá la dimensión del problema, sino que también se conseguirá develar la naturaleza de las interrelaciones existentes entre las variables originales.

Existen dos formas básicas de realizar la rotación de factores: la Rotación Ortogonal y la Rotación Oblicua según que los factores rotados sigan siendo ortogonales o no. Conviene advertir que tanto en la rotación ortogonal, como en la rotación oblicua la comunalidad de cada variable no se modifica, es decir, la rotación no afecta a la bondad de ajuste de la solución factorial: aunque cambie la matriz factorial, las especificidades no cambian, y por tanto, las comunalidades permanecen inalteradas. Sin embargo, cambia la varianza explicada por cada factor, luego los nuevos factores rotados no están ordenados de acuerdo con la información que contienen, cuantificada a través de su varianza.

3.3.1 Diseño metodológico

El análisis de componentes principales de forma general:

“Supongamos que tenemos N observaciones de X_1, X_2 donde:

$$x_1 = X_1 - \bar{X}_1, x_2 = X_2 - \bar{X}_2.$$

Se desea crear dos nuevas variables C_1, C_2 llamadas componentes principales. Estas nuevas variables son función lineal de x_1, x_2 por lo tanto se pueden escribir como:

$$C_1 = a_{11}x_1 + a_{12}x_2.$$

$$C_2 = a_{21}x_1 + a_{22}x_2.$$

Las medias y las varianzas de las N observaciones para C_1, C_2 son:

$$meanC_1 = meanC_2 = 0.$$

$$VarC_1 = a_{11}^2S_1^2 + a_{12}^2S_2^2 + 2a_{11}a_{12}rS_1S_2.$$

$$VarC_2 = a_{21}^2S_1^2 + a_{22}^2S_2^2 + 2a_{21}a_{22}rS_1S_2.$$

Donde:

$$S_i^2 = VarX_i.$$

Los a_{ij}^2 son los coeficientes¹¹ se escogen satisfactoriamente por tres requisitos:

$VarC_1$ es el más largo posible.

Las N valores de C_1 y C_2 son incorrelacionados

$$a_{11}^2 + a_{12}^2 = a_{21}^2 + a_{22}^2 = 1$$

Este modelo puede ser aplicado para la generación de un índice de nivel de satisfacción de servicios médicos de la UMAE; sin embargo, es necesario encontrar las variables que puedan ser incluidas en el modelo. [Afifif, May & Clark, 2012].

3.3.2 Hipótesis y variables

Los indicadores construidos presentan una distribución normal y son variables independientes.

Hipótesis 1. La prueba KMO debe ser mayor a 0.75 para poder la construcción del índice de victimización por medio del análisis de componentes principales.

Hipótesis 2. La medida de adecuación muestral individual para cada una de las variables construidas debe ser mayor a 0.5, es decir, cada una de las variables deberá ser compatible con el modelo del índice de victimización.

Hipótesis 3. Se rechaza la hipótesis nula en el *test* Bartlett, es decir, que la matriz de correlaciones es significativamente diferente de una matriz identidad y las variables están incorrelacionadas.

Hipótesis 4. Las primeras cuatro componentes explican el modelo del índice de victimización muy cercano al 80% de la varianza contenida en las variables originales.

Hipótesis 5. Se puede expresar mediante la combinación lineal una ecuación del índice de victimización basado en los coeficientes de las componentes principales.

Por otro lado, para la construcción del índice de victimización y percepción sobre seguridad pública deberán seleccionarse las variables a analizar y los elementos en la población en donde serán observadas.

¹¹ La solución matemática para los coeficientes fue originalmente derivada por Hottelling (1993).

Cuadro 3.1
Descripción de variables

Variable	Mnemónico	Descripción
Tasa de prevalencia delictiva por entidad federativa por cada cien mil habitantes, según sexo de la víctima, 2020.	F1_Tasa_de_prevalencia	La tasa se calcula dividiendo el total de víctimas en la entidad federativa entre la población de 18 años y más residente en ésta, multiplicada por 100 000 habitantes.
Tasa de prevalencia delictiva por entidad federativa por cada cien mil habitantes, según tipo de delito, 2020. Robo total o parcia de vehículos.	F2_Robo_total_o_parcial_de_vehiculos	La tasa se calcula dividiendo el total de víctimas en la entidad federativa entre la población de 18 años y más residente en ésta, multiplicada por 100 000 habitantes.
Tasa de prevalencia delictiva por entidad federativa por cada cien mil habitantes, según tipo de delito, 2020. Fraude	F2_Fraude	La tasa se calcula dividiendo el total de víctimas en la entidad federativa entre la población de 18 años y más residente en ésta, multiplicada por 100 000 habitantes.
Tasa de incidencia delictiva por entidad federativa de ocurrencia por cada cien mil habitantes y los cinco delitos más frecuentes 2020. Fraude.	F3_Fraude	La tasa se calcula dividiendo el total de delitos en la entidad federativa de ocurrencia entre la población de 18 años y más multiplicada por 100 000 habitantes.
Tasa de incidencia delictiva por entidad federativa de ocurrencia por cada cien mil habitantes y los cinco delitos más frecuentes 2020. Extorsión	F3_Extorsion	La tasa se calcula dividiendo el total de delitos en la entidad federativa de ocurrencia entre la población de 18 años y más multiplicada por 100 000 habitantes.
Víctimas por entidad federativa, según lugar de victimización 2020	F4_Relativos_Victimias	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de 18 años y más por entidad federativa, según condición de victimización 2020	F5_Relativos_No_Victimias	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Se consumen Alcohol en la calle.	F6_Se_consume_alcohol_en_la_calle	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su	F6_Se_consume_droga	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.

Variable	Mnemónico	Descripción
vivienda marzo y abril de 2021. Se consumen Droga.		
Población de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Existen robos o asaltos frecuentes.	F6_Existen_robos_o_a saltos_frecuentes	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de hombres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Consumo de alcohol en la calle	F7_Se_consume_alcohol_en_la_calle	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de hombres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Se consume Droga	F7_Se_consume_droga	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de hombres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Existen robos o asaltos Frecuente	F7_Existen_robos_o_a saltos_frecuentes	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de mujeres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Se consume alcohol en la calle	F8_Se_consume_alcohol_en_la_calle	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de mujeres de 18 años y más por entidad federativa y las diez conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Se consume Droga	F8_Se_consume_droga	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.
Población de mujeres de 18 años y más por entidad federativa y las diez	F8_Existen_robos_o_a saltos_frecuentes	Se estimó tomando como referencia el coeficiente de variación CV (%). Requiere un uso

Variable	Mnemónico	Descripción
conductas delictivas o antisociales que se reportan con mayor frecuencia, según conocimiento de la ocurrencia de las mismas en los alrededores de su vivienda marzo y abril de 2021. Existen robos o asaltos frecuentes.		cauteloso de la estimación en el que se analicen las causas de la alta variabilidad y se consideren otros indicadores de precisión y confiabilidad, como el intervalo de confianza.

Fuente: Elaboración propia

3.3.3 Criterios y estandarización de variables

Es importante contar con una idea más o menos clara de cuáles son los factores comunes para medir y elegir las variables de acuerdo a los mismos. La muestra debe ser representativa de la población objeto de estudio y del mayor tamaño posible. Los resultados del análisis no tienen por qué ser invariantes a cambios de origen y escala por lo que si las unidades de medida de las variables no son comparables se deben estandarizar los datos antes de realizar el análisis.

Después de obtener la matriz de componentes, se definen las puntuaciones factoriales mediante una estimación para cada uno de los sujetos en cada factor extraído, con el objeto de valorar la situación que tiene ese sujeto frente a las variables construidas (factores).

El cálculo de las puntuaciones factoriales son las puntuaciones que tienen los componentes principales para cada uno de los sujetos de la muestra, dichas puntuaciones se calculan mediante la expresión:

$$X_{ij} = a_{i1} \cdot Z_{1j} + \dots + a_{ip} \cdot Z_{pj} = \sum_{s=1}^k a_{is} \cdot Z_{sk}$$

Donde:

a_{is} – son los coeficientes

Z_{sk} – son los valores estandarizados

Al utilizar el método de extracción de factores, las puntuaciones obtenidas se calculan a partir de la solución factorial, esta solución es ortogonal¹² por lo que las puntuaciones también lo son.

3.3.4 Matriz de correlación

La matriz de correlación entre las variables observadas es comúnmente un arreglo de forma cuadrática:

$$R_p = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

Donde $r_{ij} = \frac{s_{ij}}{(s_i s_j)^{\frac{1}{2}}}$

¹² Independencia matemática de ejes factoriales entre sí (i.e., ángulos rectos o ángulos de 90 grados).

$S_i, i = (1, \dots, p)$ son las varianzas de cada una de las variables observadas

$S_{ij}, i = (1, \dots, p), j = (1, \dots, p)$ son las covarianzas de par en par de las variables observadas.

Cuadro 3.2
Matriz de Correlación

	F2_Robo_t otal_o_par cial_de_ve hiculos	F2_Fraude	F3_Fraud e	F3_Extor sion	F4_Relativo s_Victim as	F5_Relati vos_No_ Victimas	F6_Se_co nsume_al cohol_en _la_calle	F6_Se_co nsume_d roga	F6_Exist e_n_robos_ o_asaltos _frecuent es	F7_Se_con sume_alco hol_en_la_ calle	F7_Se_ con sume_dr oga	F7_Exist en_robos_ o_ asaltos_fre cuentes	F8_Se_ con sume_alc ohol_e n_la_ca lle	F8_Se_co nsume_d roga	F8_Exist e_n_robos_ o_asaltos _frecuent es
F2_Robo_total_o_parcial_de_vehiculos	1.000	-0.101	0.101	-0.321	0.415	-0.427	-0.288	0.400	0.295	-0.259	0.444	0.267	-0.309	0.455	0.304
F2_Fraude	-0.101	1.000	0.873	-0.307	-0.037	0.034	-0.023	0.117	0.117	-0.067	0.029	0.077	0.014	0.044	0.164
F3_Fraude	0.101	0.873	1.000	-0.459	-0.035	0.031	-0.205	0.109	0.164	-0.236	0.015	0.104	-0.177	0.058	0.226
F3_Extorsion	-0.321	-0.307	-0.459	1.000	-0.387	0.398	0.085	-0.302	-0.563	0.107	-0.256	-0.500	0.064	-0.264	-0.600
F4_Relativos_Victim as	0.415	-0.037	-0.035	-0.387	1.000	-0.999	0.316	0.610	0.842	0.332	0.688	0.838	0.298	0.634	0.825
F5_Relativos_No_Victima s	-0.427	0.034	0.031	0.398	-0.999	1.000	-0.307	-0.609	-0.844	-0.323	-0.689	-0.841	-0.289	-0.636	-0.826
F6_Se_consume_alcohol _en_la_calle	-0.288	-0.023	-0.205	0.085	0.316	-0.307	1.000	0.489	0.427	0.991	0.513	0.427	0.994	0.491	0.390
F6_Se_consume_droga	0.400	0.117	0.109	-0.302	0.610	-0.609	0.489	1.000	0.643	0.502	0.947	0.605	0.471	0.952	0.647
F6_Existen_robos_o_asaltos_frecuentes	0.295	0.117	0.164	-0.563	0.842	-0.844	0.427	0.643	1.000	0.398	0.667	0.976	0.445	0.617	0.991
F7_Se_consume_alcohol _en_la_calle	-0.259	-0.067	-0.236	0.107	0.332	-0.323	0.991	0.502	0.398	1.000	0.531	0.403	0.969	0.506	0.358
F7_Se_consume_droga	0.444	0.029	0.015	-0.256	0.688	-0.689	0.513	0.947	0.667	0.531	1.000	0.644	0.491	0.961	0.657
F7_Existen_robos_o_asaltos_frecuentes	0.267	0.077	0.104	-0.500	0.838	-0.841	0.427	0.605	0.976	0.403	0.644	1.000	0.440	0.582	0.948
F8_Se_consume_alcohol _en_la_calle	-0.309	0.014	-0.177	0.064	0.298	-0.289	0.994	0.471	0.445	0.969	0.491	0.440	1.000	0.471	0.412
F8_Se_consume_droga	0.455	0.044	0.058	-0.264	0.634	-0.636	0.491	0.952	0.617	0.506	0.961	0.582	0.471	1.000	0.612
F8_Existen_robos_o_asaltos_frecuentes	0.304	0.164	0.226	-0.600	0.825	-0.826	0.390	0.647	0.991	0.358	0.657	0.948	0.412	0.612	1.000

Determinante = = 1.971E-18

Fuente: Elaboración propia a partir de la aplicación del modelo factorial.

Como podemos observar, los elementos de la diagonal presentan una correlación igual a 1, mientras que el resto de los elementos sus correlaciones son menores a 1. El valor numérico de las correlaciones oscila entre [-1,1], cuando la correlación relativamente cercana a -1 o 1 significa que existe una correlación entre ambas variables, mientras que, si da cero, significa que ambas variables no tienen correlación alguna.

Adicionalmente otra característica que presenta una matriz de correlación es su simetría, es decir, si se divide la matriz por la diagonal se observa que los elementos del triángulo inferior son igual a los elementos del triángulo superior.

Como se puede observar el determinante de la matriz de correlación es igual 0.009, al ser un valor que es cercano a 0, indica la conveniencia del uso del análisis de factores.

3.3.5 Factores en la solución

El coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto. Si las variables comparten factores comunes, el coeficiente de correlación parcial debe ser bajo entre pares de variables, puesto que se eliminan los efectos lineales de las otras variables.

Las correlaciones parciales son estimaciones de las correlaciones entre los factores únicos y deberían ser próximos a cero cuando el análisis de componentes principales es adecuado, ya que, estos factores se supone que están incorrelacionados entre sí. Por lo tanto, si existe un número elevado de coeficientes de este tipo distintos de cero es señal de que las hipótesis del modelo factorial no son compatibles con los datos.

La medida de adecuación de la muestra *KMO* propuesta por Kaiser, Meyer y Olkin, la cual está dada por:

$$KMO = \frac{\sum_{j \neq i} \sum_{i \neq j} r_{ij}^2}{\sum_{j \neq i} \sum_{i \neq j} r_{ij}^2 + \sum_{j \neq i} \sum_{i \neq j} r_{ij(p)}^2}$$

Donde $r_{ij(p)}^2$ es el coeficiente de correlación parcial entre las variables X_i, X_j eliminando la influencia del resto de las variables.

Es importante comentar que *KMO* puede tomar valores entre [0,1], el cual es utilizado para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial de forma que, cuanto más pequeño sea el valor, mayor es el valor de los coeficientes de correlación parcial $r_{ij(p)}^2$ y por lo tanto menos deseable es realizar un análisis de componentes principales.

El índice se puede interpretar como la potencialidad del análisis componentes principales para extraer la variabilidad de las variables originales. Por esta razón, Kaiser, Mayer y Olkin aconsejan los siguientes criterios para establecer la idea de realizar el análisis de componentes principales:

$KMO \geq 0.75$ – *Es excelente*

$0.75 > KMO \geq 0.5$ – *Es aceptable*

$KMO < 0.5$ – *Es inaceptable*

Respecto a la medida de KMO calculada para la generación del índice de victimización y percepción sobre seguridad pública, se obtuvo un resultado de **0.731** lo que corresponde al rango de *aceptable* por lo que es viable realizar el análisis de componentes principales y se cumple la hipótesis 1 establecida con anterioridad.

Cuadro 3.3
Prueba de KMO y Barlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo

Fuente: Elaboración propia a partir de la aplicación del modelo factorial.

Adicionalmente estamos en posibilidades de calcular una medida de adecuación muestral (MSA_i) para cada variable, en esta prueba se requieren de los coeficientes de las variables que se desean comprobar, mediante:

$$MSA_i = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} r_{ij(p)}^2}$$

Donde $i = 1, \dots, p$

La matriz de correlación anti-imagen es la matriz de las correlaciones parciales, en la diagonal se encuentran las MSA_i de cada una de las variables y el resto de los valores de la matriz es la correlación parcial del resto de las variables.

Por cada uno de los 16 indicadores planteados, se obtiene una medida de adecuación muestral individual mayor a 0.5 ($MSA_i \geq 0.5$). De lo anterior, se considera que se tiene un buen nivel de adecuación para considerar cada una de las variables de manera independiente dentro del modelo de análisis de componentes principales para la construcción del índice de victimización y percepción sobre seguridad pública.

En el siguiente cuadro se podrán observar los valores de la diagonal de la matriz de correlación anti-imagen (MSA_i de cada una de las variables en el modelo), por lo que se aprueba la hipótesis 2.

Cuadro 3.4
Matriz de Correlación anti-imagen

	F2_Robo_total_o_parcial_de_vehiculos	F2_Fraude	F3_Fraude	F3_Extorsion	F4_Relativos_Victimas	F5_Relativos_No_Victimas	F6_Se_consume_alcohol_en_la_calle	F6_Se_consume_droga	F6_Existen_robos_o_asaltos_frecuentes	F7_Se_consume_alcohol_en_la_calle	F7_Se_consume_droga	F7_Existen_robos_o_asaltos_frecuentes	F8_Se_consume_alcohol_en_la_calle	F8_Se_consume_droga	F8_Existen_robos_o_asaltos_frecuentes
F2_Robo_total_o_parcial_de_vehiculos	0.327	0.052	-0.039	0.027	0.005	0.005	-0.001	-0.004	-0.005	0.001	-0.022	0.014	0.001	-0.003	0.008
F2_Fraude	0.052	0.100	-0.084	0.013	0.003	0.004	0.000	-0.028	0.000	0.000	-0.013	-0.001	0.000	0.028	0.006
F3_Fraude	-0.039	-0.084	0.083	0.002	-0.002	-0.003	0.000	0.023	0.001	0.000	0.013	-0.001	0.000	-0.026	-0.008
F3_Extorsion	0.027	0.013	0.002	0.393	-0.005	-0.004	-0.001	-0.006	0.000	0.001	-0.014	-0.007	0.001	0.011	0.011
F4_Relativos_Victimas	0.005	0.003	-0.002	-0.005	0.002	0.002	-1.438E-05	-0.002	0.000	3.191E-05	-0.001	0.000	2.523E-05	0.002	0.000
F5_Relativos_No_Victimas	0.005	0.004	-0.003	-0.004	0.002	0.002	-1.892E-05	-0.002	0.000	4.289E-05	-0.001	0.000	3.253E-05	0.002	0.000
F6_Se_consume_alcohol_en_la_calle	-0.001	0.000	0.000	-0.001	-1.438E-05	-1.892E-05	1.297E-05	0.000	4.714E-05	-2.783E-05	-4.305E-06	-8.447E-05	-2.384E-05	-9.484E-05	-8.877E-05
F6_Se_consume_droga	-0.004	-0.028	0.023	-0.006	-0.002	-0.002	0.000	0.050	0.002	0.000	-0.013	-0.002	0.000	-0.026	-0.006
F6_Existen_robos_o_asaltos_frecuentes	-0.005	0.000	0.001	0.000	0.000	0.000	4.714E-05	0.002	0.004	0.000	0.000	-0.009	-8.762E-05	-0.001	-0.005
F7_Se_consume_alcohol_en_la_calle	0.001	0.000	0.000	0.001	3.191E-05	4.289E-05	-2.783E-05	0.000	0.000	5.988E-05	-7.999E-06	0.000	5.104E-05	0.000	0.000
F7_Se_consume_droga	-0.022	-0.013	0.013	-0.014	-0.001	-0.001	-4.305E-06	-0.013	0.000	-7.999E-06	0.047	-0.004	1.970E-05	-0.023	-0.001
F7_Existen_robos_o_asaltos_frecuentes	0.014	-0.001	-0.001	-0.007	0.000	0.000	-8.447E-05	-0.002	-0.009	0.000	-0.004	0.027	0.000	0.004	0.008
F8_Se_consume_alcohol_en_la_calle	0.001	0.000	0.000	0.001	2.523E-05	3.253E-05	-2.384E-05	0.000	-8.762E-05	5.104E-05	1.970E-05	0.000	4.392E-05	0.000	0.000
F8_Se_consume_droga	-0.003	0.028	-0.026	0.011	0.002	0.002	-9.484E-05	-0.026	-0.001	0.000	-0.023	0.004	0.000	0.039	0.004
F8_Existen_robos_o_asaltos_frecuentes	0.008	0.006	-0.008	0.011	0.000	0.000	-8.877E-05	-0.006	-0.005	0.000	-0.001	0.008	0.000	0.004	0.008

Fuente: Elaboración propia a partir de la aplicación del modelo factorial.

Respecto a la prueba de esfericidad de Bartlett sirve como una prueba estadística para examinar la matriz de correlación.

Sea R_ρ la matriz de correlación de las variables observadas se prueba que las correlaciones no son nulas y hay expectativas de tener éxito en la extracción de los factores. Si se confirma la hipótesis nula significa que las variables no están intercorrelacionadas, se plantea:

$$H_0: |R_\rho| = 1 \text{ o } |R_\rho| = I \text{ vs. } H_a: |R_\rho| \neq 1 \text{ o } |R_\rho| \neq I$$

El estadístico de dicho *test* está dado por:

$$d_{R_\rho} = - \left[n - 1 - \frac{1}{6}(2p + 5) \right] \log |R_\rho| = - \left[n - \frac{2p + 11}{6} \right] \sum_{j=1}^p \log(\lambda_j)$$

Donde:

n = número de individuos de la muestra
 j = 1, ..., p que son los valores propios de R_ρ

Bajo la hipótesis nula, el estadístico tiende a ser una distribución ji-cuadrada¹³ con $P(P-1)/2$ grados de libertad, es decir, si una matriz de correlación es igual a la matriz identidad entonces significa que las intercorrelaciones entre las variables son igual a cero.

Si con el *test* Bartlett se obtiene valores altos de χ^2 , esto significa que hay variables con correlaciones altas, por lo tanto, la hipótesis nula se rechaza. En caso de no rechazar la hipótesis nula significa que las variables no están intercorrelacionadas y en este supuesto debería de reconsiderarse la aplicación de un análisis de componentes principales.

El resultado que arrojo al calcular el índice de satisfacción de servicios médicos de la UMAE, la prueba de esfericidad de Bartlett dio un valor **1025.989** y significancia 0.000, lo cual es significativo y, por lo tanto, se rechaza la hipótesis nula, es decir, que la matriz de correlación sea igual a una matriz de identidad, las variables están incorrelacionadas y es un hecho en tener éxito en la extracción de los factores. Por esta razón, se aprueba la hipótesis 3 establecida anteriormente.

Cuadro 3.5
Prueba de esfericidad de Bartlett

Aprox. Chi-cuadrado	1025.989
gl	105
Sig.	0.000

Fuente: Elaboración propia a partir de la aplicación del modelo factorial.

¹³ Distribución de probabilidad utilizada para probar el valor de la varianza o desviaciones estándar de una población, la cual es identificada por el parámetro, es decir, el número de grados de libertad.

3.4 Construcción del índice de victimización

Al utilizar el método de extracción de factores, las puntuaciones obtenidas se calculan a partir de la solución factorial, esta solución es *ortogonal*¹⁴ por lo que las puntuaciones también lo son. Mediante la matriz de coeficientes para el cálculo de las puntuaciones factoriales es posible construir una ecuación lineal de la primera componente para construir el *índice de victimización y percepción sobre seguridad pública* por cada una de las observaciones.

Mediante la matriz de coeficientes para el cálculo de las puntuaciones factoriales es posible construir una ecuación lineal de la primera componente, que nos da el nivel de victimización para cada una de las observaciones.

$$I_{victimización} = a_1Z_{I_1} + a_2Z_{I_2} + a_3Z_{I_3} + \dots + a_{n-1}Z_{I_{n-1}} + a_nZ_{I_n}$$

Sustituyendo Z_{I_i} que son los valores estandarizados, se tiene que:

$$I_{victimización} = a_1 \frac{(x_{I_1} - \bar{X}_{V_1})}{\sigma_{V_1}} + a_2 \frac{(x_{I_2} - \bar{X}_{V_2})}{\sigma_{V_2}} + a_3 \frac{(x_{I_3} - \bar{X}_{V_3})}{\sigma_{V_3}} + \dots + a_{n-1} \frac{(x_{I_{n-1}} - \bar{X}_{V_{n-1}})}{\sigma_{V_{n-1}}} + a_n \frac{(x_{I_n} - \bar{X}_{I_n})}{\sigma_{V_n}}$$

La matriz de componentes es la correlación entre las variables originales y las componentes principales, y es la clave para entender la particularidad del factor, además de esta matriz es base fundamental para la obtención del porcentaje de varianza explicada de las variables originales.

$$Var(C_i) = l_1^2 + l_2^2 + \dots + l_p^2$$

Cuadro 3.6
Matriz de componentes

	1	2	3	4
F2_Robo_total_o_parcial_de_vehiculos	0.342	-0.568	-0.481	0.386
F2_Fraude	0.083	-0.382	0.851	0.157
F3_Fraude	0.073	-0.602	0.739	0.168
F3_Extorsion	-0.442	0.608	-0.214	0.203
F4_Relativos_Victimas	0.862	-0.186	-0.257	-0.217
F5_Relativos_No_Victimas	-0.862	0.198	0.260	0.214
F6_Se_consume_alcohol_en_la_calle	0.607	0.737	0.259	-0.020
F6_Se_consume_droga	0.849	0.008	-0.004	0.463
F6_Existen_robos_o_asaltos_frecuentes	0.916	-0.190	0.038	-0.303
F7_Se_consume_alcohol_en_la_calle	0.603	0.745	0.206	0.016
F7_Se_consume_droga	0.880	0.043	-0.110	0.416
F7_Existen_robos_o_asaltos_frecuentes	0.893	-0.152	-0.001	-0.340
F8_Se_consume_alcohol_en_la_calle	0.602	0.719	0.300	-0.050
F8_Se_consume_droga	0.846	0.029	-0.086	0.488
F8_Existen_robos_o_asaltos_frecuentes	0.902	-0.241	0.077	-0.281

Fuente: Elaboración propia a partir de la aplicación del modelo factorial.

¹⁴ Independencia matemática de ejes factoriales entre sí (i.e., ángulos rectos o ángulos de 90 grados).

Donde l_i son los elementos de la componente i en la matriz de componentes principales. Por tanto, para las cuatro primeras componentes se tiene que:

$$Var(C_1) = (0.342)^2 + (0.083)^2 + (0.073)^2 + (-0.442)^2 + (0.862)^2 + (-0.862)^2 + (0.607)^2 + (0.849)^2 + (0.916)^2 + (0.603)^2 + (0.880)^2 + (0.893)^2 + (0.602)^2 + (0.846)^2 + (0.902)^2 = 7.566122$$

$$Var(C_2) = (-0.568)^2 + (-0.382)^2 + (-0.602)^2 + (0.608)^2 + (-0.186)^2 + (0.198)^2 + (0.737)^2 + (0.008)^2 + (-0.190)^2 + (0.745)^2 + (0.043)^2 + (-0.152)^2 + (0.719)^2 + (0.029)^2 + (-0.241)^2 = 3.009614$$

$$Var(C_3) = (-0.481)^2 + (0.851)^2 + (0.739)^2 + (-0.214)^2 + (-0.257)^2 + (0.260)^2 + (0.259)^2 + (-0.004)^2 + (0.038)^2 + (0.206)^2 + (-0.110)^2 + (-0.001)^2 + (0.300)^2 + (-0.086)^2 + (0.077)^2 = 1.907531$$

$$Var(C_4) = (0.386)^2 + (0.157)^2 + (0.168)^2 + (0.203)^2 + (-0.217)^2 + (0.214)^2 + (-0.020)^2 + (0.463)^2 + (0.463)^2 + (0.016)^2 + (0.416)^2 + (-0.340)^2 + (-0.050)^2 + (0.488)^2 + (-0.281)^2 = 1.251058$$

$$Var(C_1) + Var(C_2) + Var(C_3) + Var(C_4) = \mathbf{13.734321}$$

Las comunalidades generadas al inicio del modelo, estas son siempre iguales a uno y al finalizar la extracción, las comunalidades tienden a disminuir (iguales o menores a uno).

Cuadro 3.7
Comunalidades

	Extracción
F2_Robo_total_o_parcial_de_vehiculos	0.820
F2_Fraude	0.901
F3_Fraude	0.942
F3_Extorsion	0.652
F4_Relativos_Victimas	0.892
F5_Relativos_No_Victimas	0.896
F6_Se_consume_alcohol_en_la_calle	0.978
F6_Se_consume_droga	0.935
F6_Existen_robos_o_asaltos_frecuentes	0.968
F7_Se_consume_alcohol_en_la_calle	0.962
F7_Se_consume_droga	0.961
F7_Existen_robos_o_asaltos_frecuentes	0.936
F8_Se_consume_alcohol_en_la_calle	0.971
F8_Se_consume_droga	0.962
F8_Existen_robos_o_asaltos_frecuentes	0.957

Método de extracción: análisis de componentes principales.

Por tanto:

$$Var(C_1) + Var(C_2) + Var(C_3) + Var(C_4) = 0.820 + 0.901 + 0.942 + 0.652 + 0.892 + 0.896 + 0.978 + 0.935 + 0.968 + 0.962 + 0.961 + 0.936 + 0.971 + 0.962 + 0.957 = \mathbf{13.734}$$

El criterio más utilizado para escoger el número de factores a considerar es con base en el porcentaje de varianza total explicada por cada factor, y cuando este llega a un porcentaje acumulado considerablemente alto, normalmente cerca del 80%, significa que el número de factores es suficiente. Así, se aprueba la hipótesis 4.

Cuadro 3.8
Varianza total explicada

Componente	Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	7.565	50.435	50.435	4.929	32.860	32.860
2	3.010	20.066	70.501	3.514	23.425	56.285
3	1.907	12.713	83.213	3.164	21.092	77.377
4	1.252	8.345	91.558	2.127	14.182	91.558

Método de extracción: análisis de componentes principales.

Mediante el método de rotación Varimax se obtuvo la matriz de componentes rotado.

Cuadro 3.9
Matriz de componentes rotado

	Componente			
	1	2	3	4
F2_Robo_total_o_parcial_de_vehiculos	0.280	-0.554	0.655	-0.079
F2_Fraude	0.009	0.056	0.000	0.947
F3_Fraude	0.083	-0.175	0.048	0.950
F3_Extorsion	-0.626	0.228	-0.075	-0.449
F4_Relativos_Victimas	0.854	0.092	0.371	-0.130
F5_Relativos_No_Victimas	-0.857	-0.081	-0.374	0.127
F6_Se_consume_alcohol_en_la_calle	0.182	0.950	0.193	-0.073
F6_Se_consume_droga	0.363	0.307	0.833	0.122
F6_Existen_robos_o_asaltos_frecuentes	0.911	0.235	0.261	0.118
F7_Se_consume_alcohol_en_la_calle	0.162	0.932	0.232	-0.117
F7_Se_consume_droga	0.413	0.308	0.834	0.007
F7_Existen_robos_o_asaltos_frecuentes	0.907	0.244	0.225	0.058
F8_Se_consume_alcohol_en_la_calle	0.197	0.952	0.158	-0.034
F8_Se_consume_droga	0.349	0.289	0.869	0.045
F8_Existen_robos_o_asaltos_frecuentes	0.902	0.202	0.266	0.178

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

a. La rotación ha convergido en 5 iteraciones.

Se aplica el criterio para la identificación de cargas significativas en la matriz de componentes rotado para identificar qué variables van a quedar agrupadas en cada componente. Así se pudo identificar que, en la primera componente se agrupan las variables: *Relativos_Victimas* y *Existen_robos_o_asaltos_frecuentes*; para la segunda componente las variables de *Se_consume_alcohol_en_la_calle*; y para la tercera componente se encuentran las variables de *Se_consume_droga* y para la cuarta componente las variables de *_Fraude*.

Después de obtener la matriz de componentes, se definen las puntuaciones factoriales mediante una estimación para cada uno de los sujetos en cada factor extraído, con el objeto de valorar la situación que tiene ese sujeto frente a las variables construidas (factores).

El cálculo de las puntuaciones factoriales son las puntuaciones que tienen los componentes principales para cada uno de los sujetos de la muestra, dichas puntuaciones se calculan mediante la expresión:

$$X_{ij} = a_{i1} \cdot Z_{1j} + \dots + a_{ip} \cdot Z_{pj} = \sum_{s=1}^k a_{is} \cdot Z_{sk}$$

Donde:

a_{is} – son los coeficientes

Z_{sk} – son los valores estandarizados

Al utilizar el método de extracción de factores, las puntuaciones obtenidas se calculan a partir de la solución factorial, esta solución es ortogonal¹⁵ por lo que las puntuaciones también lo son

Cuadro 3.10
Matriz de coeficiente de puntuación de componente rotado

	Componente			
	1	2	3	4
F2_Robo_total_o_parcial_de_vehiculos	-0.050	-0.260	0.346	-0.080
F2_Fraude	-0.069	0.074	0.011	0.470
F3_Fraude	-0.043	-0.011	0.037	0.453
F3_Extorsion	-0.192	0.087	0.109	-0.164
F4_Relativos_Victimas	0.219	-0.051	-0.040	-0.114
F5_Relativos_No_Victimas	-0.219	0.055	0.038	0.113
F6_Se_consume_alcohol_en_la_calle	-0.030	0.289	-0.021	0.011
F6_Se_consume_droga	-0.128	0.022	0.358	0.068
F6_Existen_robos_o_asaltos_frecuentes	0.247	0.018	-0.128	0.010
F7_Se_consume_alcohol_en_la_calle	-0.044	0.278	0.008	-0.010
F7_Se_consume_droga	-0.101	0.013	0.342	0.008
F7_Existen_robos_o_asaltos_frecuentes	0.259	0.021	-0.150	-0.019
F8_Se_consume_alcohol_en_la_calle	-0.017	0.294	-0.045	0.028
F8_Se_consume_droga	-0.137	0.009	0.383	0.031
F8_Existen_robos_o_asaltos_frecuentes	0.239	0.011	-0.118	0.039

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

Mediante la matriz de coeficientes para el cálculo de las puntuaciones factoriales es posible construir una ecuación lineal de la segunda componente, que es el índice de consume de alcohol y drogas en la calle para cada una de las observaciones, por lo tanto se toman las puntuaciones de la segunda componente que se muestran en el cuadro 3.10 y se aplican en la ecuación lineal del apartado 3.4.

$$I_{victimizacion} = -0.260Z_{I_1} + 0.074Z_{I_2} - 0.011Z_{I_3} + 0.087Z_{I_4} - 0.051Z_{I_5} - 0.055Z_{I_6} + 0.289Z_{I_7} + 0.022Z_{I_8} + 0.018Z_{I_9} + 0.278Z_{I_{10}} + 0.013Z_{I_{11}} + 0.021Z_{I_{12}} + 0.294Z_{I_{13}} + 0.009Z_{I_{14}} + 0.011Z_{I_{15}}$$

¹⁵ Independencia matemática de ejes factoriales entre sí (i.e., ángulos rectos o ángulos de 90 grados).

Sustituyendo Z_{I_i} que son los valores estandarizados, se tiene que:

$$I_{victimizacion} = -0.260 \frac{(x_{I_1} - \bar{X}_{I_1})}{\sigma_{I_1}} + 0.074 \frac{(x_{I_2} - \bar{X}_{I_2})}{\sigma_{I_2}} - 0.011 \frac{(x_{I_3} - \bar{X}_{I_3})}{\sigma_{I_3}} + 0.087 \frac{(x_{I_4} - \bar{X}_{I_4})}{\sigma_{I_4}} \\ - 0.051 \frac{(x_{I_5} - \bar{X}_{I_5})}{\sigma_{I_5}} + 0.055 \frac{(x_{I_6} - \bar{X}_{I_6})}{\sigma_{I_6}} + 0.289 \frac{(x_{I_7} - \bar{X}_{I_7})}{\sigma_{I_7}} + 0.222 \frac{(x_{I_8} - \bar{X}_{I_8})}{\sigma_{I_8}} \\ + 0.018 \frac{(x_{I_9} - \bar{X}_{I_9})}{\sigma_{I_9}} + 0.0278 \frac{(x_{I_{10}} - \bar{X}_{I_{10}})}{\sigma_{I_{10}}} + 0.013 \frac{(x_{I_{11}} - \bar{X}_{I_{11}})}{\sigma_{I_{101}}} + 0.021 \frac{(x_{I_{12}} - \bar{X}_{I_{12}})}{\sigma_{I_{12}}} \\ + 0.294 \frac{(x_{I_{13}} - \bar{X}_{I_{13}})}{\sigma_{I_{13}}} + 0.009 \frac{(x_{I_{14}} - \bar{X}_{I_{14}})}{\sigma_{I_{14}}} + 0.011 \frac{(x_{I_{15}} - \bar{X}_{I_{15}})}{\sigma_{I_{15}}}$$

Se conocen las medias y varianzas de los 15 indicadores

Cuadro 3.11
Resultados de medias, varianzas y desviaciones estándar de los 15 indicadores

	F2_Robo total o parcial de vehículo	F2_Fraude	F3_Fraude	F3_Extorsión	F4_Relativos Víctimas	F5_Relativos No Víctimas	F6_Se consume alcohol en la calle	F6_Se consume droga
Media	0.27825	0.20823	0.20112	0.18283	0.21835	0.77873	0.64508	0.47682
Varianza	0.00611	0.00140	0.00158	0.00494	0.00230	0.00233	0.00725	0.00948
Desviación estándar	0.07819	0.03746	0.03974	0.07028	0.04801	0.04822	0.08515	0.09739

	F6_Existen robos o asaltos frecuentes	F7_Se consume alcohol en la calle	F7_Se consume droga	F7_Existen robos o asaltos frecuentes	F8_Se consume alcohol en la calle	F8_Se consume droga	F8_Existen robos o asaltos frecuentes
Media	0.40909	0.64825	0.47208	0.38056	0.64209	0.47182	0.43143
Varianza	0.01155	0.00692	0.00836	0.01280	0.00779	0.00858	0.01192
Desviación estándar	0.10748	0.08322	0.09142	0.11312	0.08828	0.09263	0.10917

Sustituyendo \bar{X}_{I_i} y σ_{I_i} respectivamente, se obtiene:

$$I_{victimizacion} = -0.260 \frac{(x_{I_1} - 0.27825)}{0.00611} + 0.074 \frac{(x_{I_2} - 0.20823)}{0.00140} - 0.011 \frac{(x_{I_3} - 0.20112)}{0.00158} \\ + 0.087 \frac{(x_{I_4} - 0.18283)}{0.00494} - 0.051 \frac{(x_{I_5} - 0.21835)}{0.00230} + 0.055 \frac{(x_{I_6} - 0.77873)}{0.00233} \\ + 0.289 \frac{(x_{I_7} - 0.64508)}{0.00725} + 0.222 \frac{(x_{I_8} - 0.47682)}{0.00948} + 0.018 \frac{(x_{I_9} - 0.40909)}{0.01155} \\ + 0.0278 \frac{(x_{I_{10}} - 0.64825)}{0.00692} + 0.013 \frac{(x_{I_{11}} - 0.47208)}{0.00836} + 0.021 \frac{(x_{I_{12}} - 0.38056)}{0.01280} \\ + 0.294 \frac{(x_{I_{13}} - 0.64209)}{0.00779} + 0.009 \frac{(x_{I_{14}} - 0.47182)}{0.00858} + 0.011 \frac{(x_{I_{15}} - 0.43143)}{0.01192}$$

Simplificando, mediante la ecuación lineal para el cálculo del índice de victimización se aprueba la hipótesis 5. La ecuación está dada por:

$$I_{\text{victimizacion}} = -0.260(x_{I_1} - 45.5156) + 0.074(x_{I_2} - 148.42341) - 0.11(x_{I_3} - 127.3291) + 0.087(x_{I_4} - 37.0128) \\ - 0.051(x_{I_5} - 94.7412) + 0.55(x_{I_6} - 334.9312) + 0.289(x_{I_7} - 88.9632) + 0.222(x_{I_8} - 50.2735) \\ + 0.018(x_{I_9} - 35.4145) + 0.0278(x_{I_{10}} - 93.6121) + 0.013(x_{I_{11}} - 56.4818) + 0.021(x_{I_{12}} - 29.7429) \\ + 0.0294(x_{I_{13}} - 82.39700) + 0.009(x_{I_{14}} - 54.9864) + 0.011(x_{I_{15}} - 36.2011))$$

Este índice es útil porque permite resumir en pocos factores o componentes todas las variables observables de la ENVIPE. En 15 factores se obtienen nuevas variables que pueden explicar con más éxito alguna estructura al resumir las variables y obtener los puntajes de los factores, esos puntajes forman nuevas variables.

Ahora la pregunta sería *¿Qué se puede hacer con las nuevas variables que se encontraron?* La respuesta es: utilizarlas en el análisis de regresión múltiple, en el análisis discriminante, etc., las cuales son técnicas dependientes y que requieren de 2 o más variables independientes métricas que no estén correlacionadas y que afecten la variable que se desea pronosticar; por lo que si las variables independientes tienen problemas de multicolinealidad se recomienda que se ortogonalicen por medio de esta ecuación que es resultado del análisis de componentes principales.

3.5 Conclusión del capítulo

La mayoría de las variables extraídas de la ENVIPE 2020 fueron posibles agruparlas para aplicar como técnica estadística el método multivariado de componentes principales. Esta técnica permitió identificar un número relativamente pequeño de factores que pudieron ser usados a su vez para representar en forma resumida la información contenida en un conjunto numeroso de variables interrelacionada que permitieron construir un índice de victimización.

Entre las fortalezas del método para construir el índice están en que se tiene acceso a las bases de datos públicas de la ENVIPE 2020 para extraer la información con fácil accesibilidad y compatibilidad con paquetes estadísticos de uso común; además que es una metodología utilizada por diferentes organismos gubernamentales y empresas privadas. Los resultados pueden ser extrapolados a nivel de entidad federativa.

En sentido opuesto, las debilidades en la construcción del índice de violencia es que los resultados publicados de manera oficial no pudieran representar la problemática real del país y la metodología de cálculo pudiera no ser comprensiva para los especialistas cuya formación profesional no sea científica y más bien sea humanística. Finalmente, las amenazas estarían circunscritas en la interpretación y consistencia de los resultados y que metodología solo se utilice para fines académicos y no para investigación aplicada o instrumentación de políticas públicas.

Conclusiones

Los estudios de opinión pública, en este caso, la aplicación de encuestas «cara a cara» a la población de 18 años y más sobre la seguridad del lugar donde viven (colonia o localidad, municipio, entidad federativa) y sobre los lugares en donde realizan sus actividades cotidianas (trabajo, escuela, mercado, transporte público, entre otros) han demostrado ser una herramienta útil y confiable para recolectar información de la percepción sobre el fenómeno de la victimización delictiva; así como la percepción social respecto de la seguridad pública y el desempeño de las autoridades.

La ventaja de la aplicación del análisis factorial, ocurre desde la preparación de la información, se inicia con la simplificación de la base de datos, es decir, con la reducción de variables que no predominan o que se intuye no serán necesarias para la construcción del índice de victimización y percepción sobre seguridad pública; esta técnica estadística de reducción de datos facilita el manejo de grandes volúmenes de información y su análisis.

El análisis descriptivo de los datos recopilados en la encuesta permite la selección de las variables que se consideran fundamentadas por los aspectos teóricos del problema, aquellos denominados factores. Posteriormente, se normalizan y estandarizan los factores y con el uso del software estadístico se profundiza en el análisis matemático con la aplicación de diversas pruebas, como el test de esfericidad de Barlett y la prueba de KMO. A través de estas pruebas se hace la exploración y confirmación de las hipótesis establecidas. Destaca que, con los factores elegidos el valor obtenido del índice KMO (0.731), lo ubicó en el rango de excelente y se pudo proceder a la construcción del índice de índice de victimización y percepción sobre seguridad pública por medio del análisis de componentes principales. Esto significa que con un número relativamente pequeño de variables (16) se representó la información contenida en un conjunto amplio de variables interrelacionadas.

Bibliografía

- 📖 Afifi, A., May S. & Clark V. Practical multivariate analysis. CRC Press.USA, 2012.
- 📖 Cuadras Carles M. Nuevos Métodos de análisis multivariante, CMC Editions, España 2014
- 📖 De la Garza García Jorge et aut. Análisis estadístico multivariante. Un enfoque teórico y práctico. Editorial Mc Graw Hill. México. 2013
- 📖 Dillon William & Goldstein Matthew. Multivariate analysis. Methods and applications. John Wiley & Sons. USA, 1984.
- 📖 Hari J., Anderson R., Tatham R. & Black W Multivariate data analysis with readings.. Prentice Hall. New Jersey, 1995. Cuadras, Carles M. Nuevos Métodos de Análisis Multivariante. España, 2007
- 📖 INEGI. Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública, 2020. México.
- 📖 Inegi. Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública, 2020. México. <https://www.inegi.org.mx/programas/envipe/2020/>
- 📖 Johnson, R.A. & Wichern, D.W. Applied multivariate statistical analysis, 1988.
- 📖 Maronna Ricardo A., Probabilidad y Estadística Elementales para estudiantes de ciencias, Facultad de ciencias exactas –Universidad Nacional de La Plata-, 1995
- 📖 Menéndez Eduardo L. Violencias en México: las explicaciones y las ausencias. Centro de Investigaciones y Estudios Superiores en Antropología Social. México 2011.
- 📖 Oviedo Jorge Mauricio, Matriz de varianzas vs matriz de correlaciones en el Análisis de componentes principales: Un enfoque analítico, McGraw-Hill, Madrid; 2002
- 📖 Rencher Alvin C., Methods of multivariate analysis, wiley-interscience, 2002
- 📖 Santaló Luis A., Probabilidad e inferencia estadística, Programa Regional de Desarrollo Científico y Tecnológico. Departamento de Asuntos científicos, Secretaría General de la Organización de los Estados Americanos, Washington, D.C., 1975