



Instituto Nacional  
de Estadística

---

# Guía técnica para el manejo de bases de datos

---



Guatemala, octubre 2024



## Contenido

Introducción .....	i
Objetivo.....	1
Contenido temático de la encuesta .....	1
Acceso a las bases de datos .....	2
Estructura de las bases de datos.....	5
Uso de factores de expansión.....	7
Precisión de las estimaciones.....	8
Especificaciones del diseño estadístico .....	9
Contenido de las bases de datos.....	9
Base de agregado de consumo.....	10
Base de personas.....	10
Base de hogares.....	12
Base de fuentes .....	12
Base de donaciones.....	13
Bases de emigración .....	13
Base de equipamiento.....	14
Base de violencia.....	14
Base de programas .....	15
Bases de gastos .....	15
Bases de alimentos.....	15
Bases de negocios .....	16
Bases de tierra .....	16
Base de unidad .....	16
Base de producción .....	17
Base de pecuarios.....	17
Base de productos.....	17
Base de instrumentos.....	17
Base de equipo.....	18

Base de instalaciones.....	18
Base de asistencia .....	18
Base de forestal.....	19
Bases de préstamos.....	19
Bases de crédito .....	20
Compatibilidad y uso en otros programas.....	20
Ejemplo del uso de las bases de datos .....	21

## Introducción

Este documento constituye una guía técnica para el procesamiento de las bases de datos de la Encuesta Nacional de Condiciones de Vida -ENCOVI 2023- y su propósito es orientar a los usuarios en el uso y manejo de estas. Con la ENCOVI 2023 se generaron bases de datos sobre las viviendas, los hogares y las personas, las cuales permiten indentificar las condiciones de vida de la población.

A lo largo de esta guía, se explican las principales características de cada base de datos, así como las herramientas necesarias para su correcta interpretación y utilización. Los usuarios deben utilizar los factores de expansión para extrapolar resultados representativos a la población y evaluar la precisión de sus estimaciones utilizando indicadores como el coeficiente de variación y los intervalos de confianza. Con esta guía, se busca motivar a los usuarios para que puedan realizar análisis profundos y efectivos, contribuyendo así al entendimiento y mejora de las condiciones de vida en Guatemala.

Considerando que las bases de datos constituyen un recurso fundamental que integra una gran cantidad de información estructurada sobre diversos temas, la ENCOVI 2023 está meticulosamente organizada en 35 bases de datos para garantizar la facilidad de acceso y la comprensión por parte de los usuarios, así como el análisis y la extracción de conocimientos.

## Objetivo

Proporcionar a los usuarios de la información los elementos necesarios para el uso adecuado y eficiente de las bases de datos de la ENCOVI 2023, mediante una descripción detallada de los aspectos clave a considerar al generar estimaciones y realizar análisis en temas como pobreza, consumo, ingresos y empleo.

## Contenido temático de la encuesta

La Encuesta Nacional de Condiciones de Vida -ENCOVI-, permite conocer las condiciones socioeconómicas de las personas y hogares guatemaltecos, proporciona información fundamental para el desarrollo de políticas públicas efectivas y el seguimiento de los avances en materia de bienestar social en el país. La edición 2023 está organizada en diecisiete capítulos, que cubren una amplia variedad de temas relevantes de la realidad económica y social del país, con un enfoque especial en la pobreza y el bienestar. A continuación, se describen los módulos temáticos que la encuesta incluye:

- I. La vivienda y el hogar
- II. Seguridad ciudadana
- III. Participación organizaciones y programas de asistencia social
- IV. Características de los miembros del hogar
- V. Salud
- VI. Educación
- VII. Capacitación para el trabajo
- VIII. Migración
- IX. Uso del tiempo
- X. Fuerza de trabajo
- XI. Otros ingresos
- XII. Uso de tecnología
- XIII. Gastos de autoconsumo
- XIV. Negocios no agropecuarios del hogar
- XV. Módulo de tenencia de la tierra
- XVI. Actividades agropecuarias
- XVII. Préstamos y compras al crédito

Es relevante mencionar que la -ENCOVI 2023- incluye un módulo específico sobre empleo, lo que posibilita la generación de indicadores precisos acerca de la situación laboral de los guatemaltecos, facilitando una comprensión más profunda del mercado laboral en el país y mantener el seguimiento conforme a la información proporcionada por los estimadores y cálculos de la Encuesta Nacional de Empleo e Ingresos -ENEI-, considerando que el INE no realizó esta última en el 2023.

## Acceso a las bases de datos

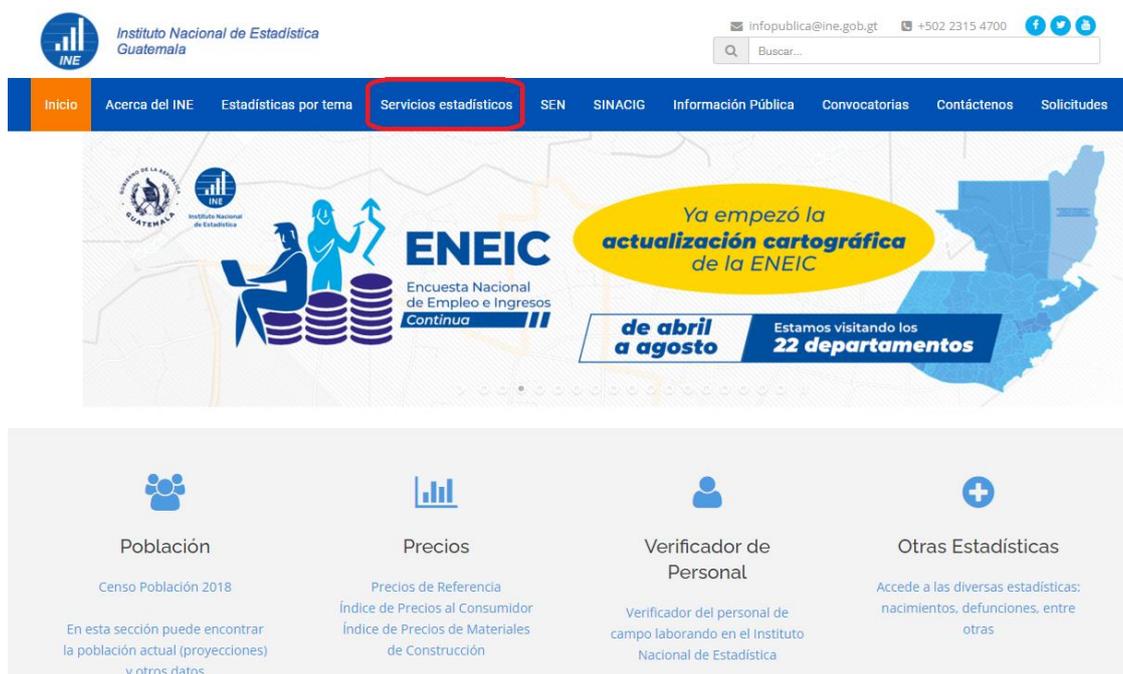
El Instituto Nacional de Estadística ha puesto a disposición en su página web las bases de datos de la ENCOVI 2023, con sus respectivos diccionarios, en dos formatos: SPSS y Excel, con el objetivo de facilitar el acceso y análisis de la información a los usuarios, permitiendo a investigadores y analistas utilizar la base de datos en el software de su preferencia. A continuación, se presentan las instrucciones detalladas para acceder y descargar las bases de datos de la encuesta desde el sitio web del INE.

### Paso 1

Acceder a la página web del instituto nacional de estadística de Guatemala mediante el siguiente enlace: <https://www.ine.gob.gt/>

### Paso 2

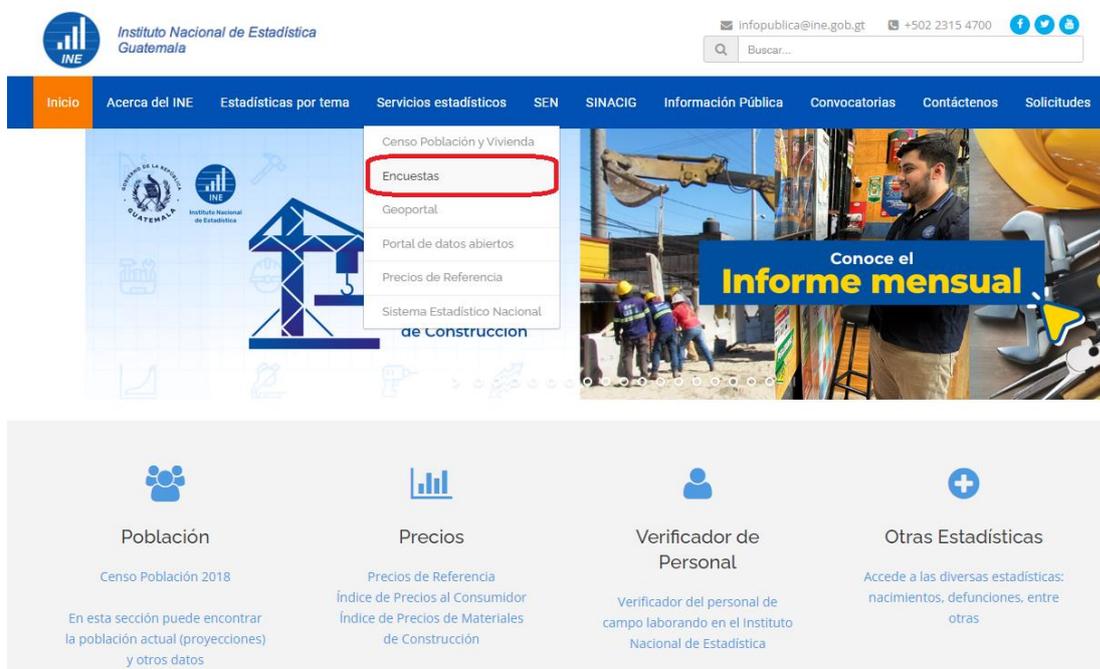
Seleccionar la pestaña servicios estadísticos en la página de inicio del sitio web.



The screenshot shows the homepage of the Instituto Nacional de Estadística (INE) of Guatemala. At the top, there is a navigation bar with the following items: Inicio, Acerca del INE, Estadísticas por tema, **Servicios estadísticos** (highlighted with a red box), SEN, SINACIG, Información Pública, Convocatorias, Contáctenos, and Solicitudes. Below the navigation bar is a large banner for the 'Encuesta Nacional de Empleo e Ingresos Continua' (ENEIC). The banner features the INE logo, a graphic of a person at a computer, and text that reads: 'Ya empezó la actualización cartográfica de la ENEIC de abril a agosto. Estamos visitando los 22 departamentos'. Below the banner is a grid of four service categories: 'Población' (Censo Población 2018), 'Precios' (Índice de Precios al Consumidor, Índice de Precios de Materiales de Construcción), 'Verificador de Personal' (Verificador del personal de campo laborando en el Instituto Nacional de Estadística), and 'Otras Estadísticas' (Accede a las diversas estadísticas: nacimientos, defunciones, entre otras).

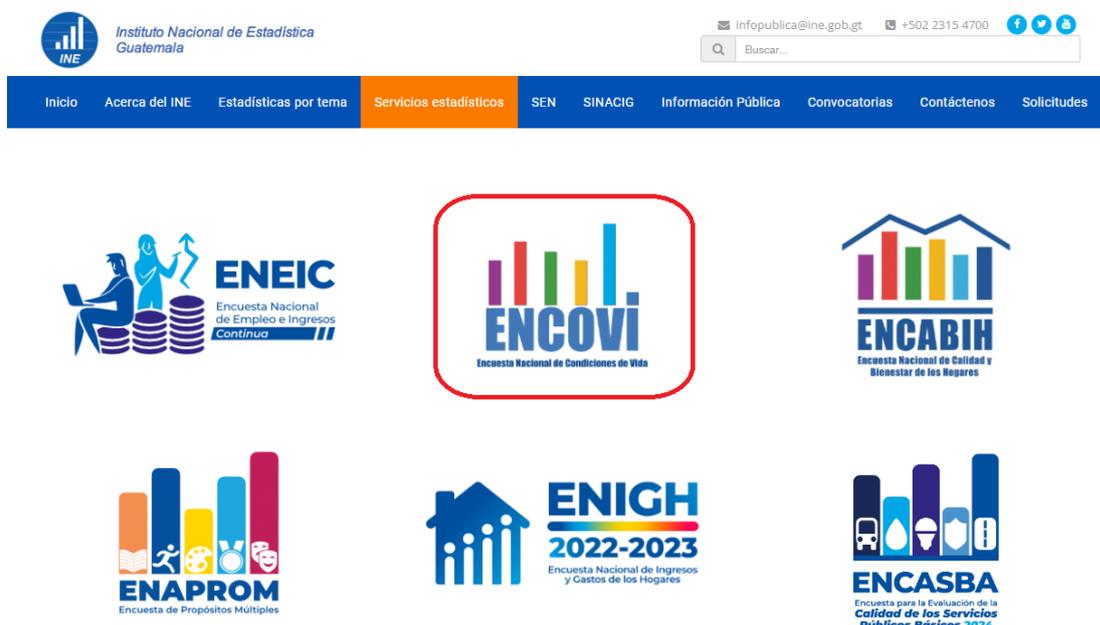
### Paso 3

Entre las opciones que se desglosan elegir la opción “Encuestas”.



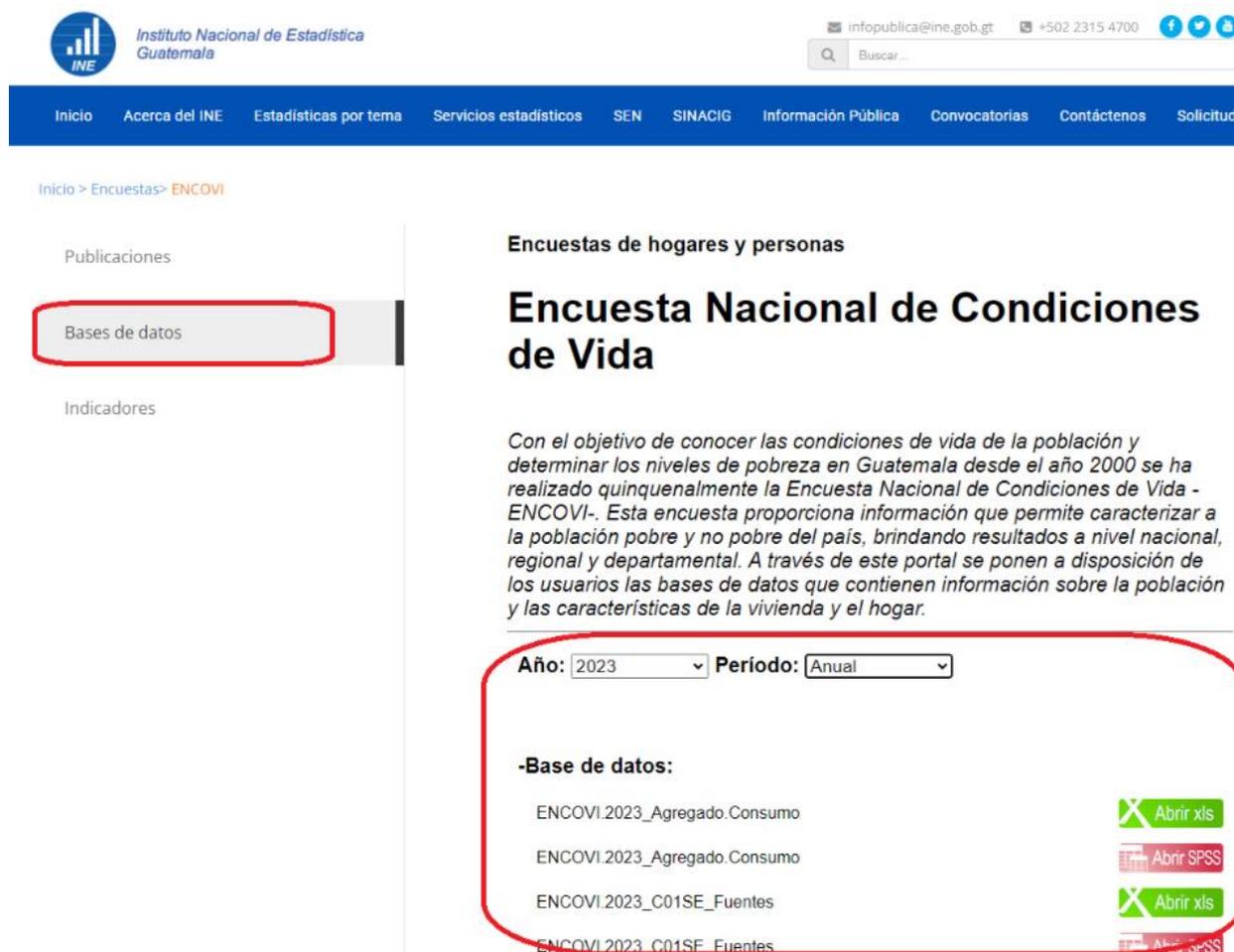
### Paso 4

En esta sección, se muestra un listado de las encuestas realizadas por el INE. Dentro de este listado, localice la ENCOVI y haga clic sobre su logotipo para continuar.



## Paso 5

A la izquierda, se verá un listado de materiales relacionados con la ENCOVI. Seleccione la opción "Bases de datos"; al hacerlo, aparecerán dos botones que permiten elegir el año y el período correspondientes.



The screenshot shows the website of the Instituto Nacional de Estadística (INE) of Guatemala. The page is titled "Encuestas de hogares y personas" and "Encuesta Nacional de Condiciones de Vida". On the left sidebar, the "Bases de datos" option is highlighted with a red box. The main content area features a description of the ENCOVI survey and a selection interface. This interface includes dropdown menus for "Año" (Year) set to 2023 and "Período" (Period) set to Anual. Below these, a section titled "-Base de datos:" lists three data files with corresponding "Abrir xls" and "Abrir SPSS" buttons, all of which are highlighted with a red box.

**Encuestas de hogares y personas**

## Encuesta Nacional de Condiciones de Vida

Con el objetivo de conocer las condiciones de vida de la población y determinar los niveles de pobreza en Guatemala desde el año 2000 se ha realizado quinquenalmente la Encuesta Nacional de Condiciones de Vida - ENCOVI-. Esta encuesta proporciona información que permite caracterizar a la población pobre y no pobre del país, brindando resultados a nivel nacional, regional y departamental. A través de este portal se ponen a disposición de los usuarios las bases de datos que contienen información sobre la población y las características de la vivienda y el hogar.

Año: 2023 Período: Anual

**-Base de datos:**

ENCOVI.2023_Agregado_Consumo	 <a href="#">Abrir xls</a>
ENCOVI.2023_Agregado_Consumo	 <a href="#">Abrir SPSS</a>
ENCOVI.2023_C01SE_Fuentes	 <a href="#">Abrir xls</a>
ENCOVI.2023_C01SE_Fuentes	 <a href="#">Abrir SPSS</a>

## Estructura de las bases de datos

Las bases de datos están diseñadas para facilitar su uso y comprensión por parte de los usuarios. La información se organiza de manera clara y sistemática: las variables se listan en las columnas, representando los diferentes aspectos o características que se desean analizar, mientras que las filas contienen las observaciones, es decir, los datos específicos de cada individuo u hogar encuestado. Esta estructura tabular permite realizar búsquedas, análisis y cruces de información de forma eficiente, adaptándose a diversas necesidades analíticas. En este sentido, la información recopilada por la encuesta se encuentra dividida en 35 bases de datos. A continuación, se detalla el listado de bases de datos de la ENCOVI 2023 disponibles en el sitio web de INE:

1. ENCOVI.2023\_Agregado.Consumo
2. ENCOVI.2023\_Personas
3. ENCOVI.2023\_Hogares
4. ENCOVI.2023\_C01SE\_Fuentes
5. ENCOVI.2023\_C01SF\_Donaciones
6. ENCOVI.2023\_C01SI\_Emigracion.1
7. ENCOVI.2023\_C01SI\_Emigracion.2
8. ENCOVI.2023\_C01SJ\_Equipamiento
9. ENCOVI.2023\_C02SA\_Violencia
10. ENCOVI.2023\_C03SA\_Programas
11. ENCOVI.2023\_C13SB\_Gastos.1
12. ENCOVI.2023\_C13SB\_Gastos.2
13. ENCOVI.2023\_C13SB\_Gastos.3
14. ENCOVI.2023\_C13SB\_Gastos.4
15. ENCOVI.2023\_C13SA\_Alimentos.1
16. ENCOVI.2023\_C13SA\_Alimentos.2
17. ENCOVI.2023\_C13SA\_Alimentos.3
18. ENCOVI.2023\_C13SA\_Alimentos.4
19. ENCOVI.2023\_C14SA\_Negocios.1
20. ENCOVI.2023\_C14SA\_Negocios.2
21. ENCOVI.2023\_C15SA\_Tierra.1
22. ENCOVI.2023\_C15SA\_Tierra.2
23. ENCOVI.2023\_C16SA\_Unidad

24. ENCOVI.2023\_C16SBC\_Produccion
25. ENCOVI.2023\_C16SD\_Pecuarios
26. ENCOVI.2023\_C16SE\_Productos
27. ENCOVI.2023\_C16SF\_Instrumentos
28. ENCOVI.2023\_C16SG\_Equipo
29. ENCOVI.2023\_C16SH\_Instalaciones
30. ENCOVI.2023\_C16SI\_Asistencia
31. ENCOVI.2023\_C16SJ\_Forestal
32. ENCOVI.2023\_C17SA\_Prestamos.1
33. ENCOVI.2023\_C17SA\_Prestamos.2
34. ENCOVI.2023\_C17SB\_Credito.1
35. ENCOVI.2023\_C17SB\_Credito.2

Es importante destacar que cada una de las bases de datos mencionadas previamente incluye variables cartográficas para la identificación de cada observación, así como una variable de caracterización de la pobreza, que identifica si la persona o el hogar es pobre extremo, pobre no extremo o no pobre, además de las variables específicas relacionadas con la temática correspondiente.

#### Variables cartográficas identificadoras:

- Departamento
- Dominio
- Área
- UPM
- No. Hogar

Dentro de las bases de datos existen registros sin clasificación en la variable de pobreza, ya que los hogares a los que pertenecen no reportaron ningún consumo, lo que impidió su categorización.

*La variable **NO\_HOGAR** es una variable clave que sirve como identificador único para cada hogar dentro de la base de datos. Su función es garantizar que cada registro de hogares sea distinto y fácilmente identificable, evitando duplicados. Esta variable es indispensable para vincular información entre diferentes bases, como la de hogares y la de personas, permitiendo realizar análisis a nivel de hogar y facilitar el cruce de datos.*

## Uso de factores de expansión

Los factores de expansión son elementos clave para realizar análisis representativos a nivel poblacional a partir de los datos recolectados en la encuesta. Cada registro en la base de datos se asocia con un factor de expansión que indica cuántas personas, hogares u otras unidades representan en la población total. Al utilizar estos factores, es posible extrapolar los resultados de la muestra a la población general, obteniendo así estimaciones más precisas de las diferentes variables analizadas.

Los factores de expansión deben aplicarse adecuadamente en los cálculos estadísticos para garantizar la validez de los resultados y su interpretación correcta, evitando sesgos y asegurando que las conclusiones reflejen de manera fiel la realidad estudiada.

### Variable **FACTOR**

Esta variable proporciona la ponderación asignada a cada uno de los hogares incluidos en la encuesta. Representa el factor de expansión que permite ajustar la representación de cada hogar según su presencia relativa en la población.

*La variable **FACTOR** se encuentra en todas las bases de datos.*

***Factor3** se emplea para calcular indicadores per cápita, siendo especialmente útil en análisis de consumo y pobreza. Esta variable está disponible únicamente en la base de datos denominada "Agregado.Consumo".*

### Variable **Factor3**

Esta variable contiene el factor de expansión ajustado según el número de personas que viven en cada hogar. La aplicación de este factor permite realizar cálculos a nivel per cápita, proporcionando una representación más precisa de los datos individuales dentro de la población.

Los diferentes factores de expansión se aplican multiplicando cada observación por su respectivo factor de expansión. Esto ajusta los cálculos estadísticos para representar correctamente la población total a partir de la muestra. Por ejemplo, al calcular promedios, totales o proporciones, cada observación se pondera según su factor.

## Precisión de las estimaciones

La precisión de las estimaciones se evalúa utilizando indicadores como el coeficiente de variación y los intervalos de confianza. El coeficiente de variación (CV) mide la dispersión relativa de las estimaciones, expresada como un porcentaje. Un CV bajo indica una mayor precisión, mientras que un CV alto señala una menor fiabilidad en las estimaciones.

Por otro lado, los intervalos de confianza ofrecen un rango dentro del cual se espera que se encuentre el valor verdadero de la población con un determinado nivel de confianza, comúnmente al 95%. Un intervalo de confianza estrecho sugiere mayor precisión, mientras que uno más amplio indica mayor incertidumbre en la estimación. Juntos, estos indicadores permiten evaluar la precisión y confiabilidad de los resultados obtenidos a partir de la muestra.

Para evaluar la precisión relativa de un indicador se calcula el coeficiente de variación aplicando la siguiente fórmula:

$$CV(\hat{\theta}) = \frac{se(\hat{\theta})}{\hat{\theta}} \times 100$$

- $\hat{\theta}$  : estimación del indicador.
- $se(\hat{\theta})$  : error estándar de la estimación.

### **Criterios sugeridos para evaluar la precisión a través del CV:**

- $\leq 20\%$  → **Indicador preciso**
- $21\% - 30\%$  → **Indicador poco preciso**
- $> 30\%$  → **Indicador impreciso**

El intervalo de confianza para un indicador estimado  $\hat{\theta}$  al nivel de confianza del 95% se expresa como:

$$IC = \left( \hat{\theta} \pm Z_{\alpha/2} * se(\hat{\theta}) \right)$$

- $\hat{\theta}$  : estimación del indicador.
- $Z_{\alpha/2}$  : valor crítico de la distribución normal estándar correspondiente al nivel de confianza deseado.
- $se(\hat{\theta})$  : error estándar de la estimación.

Además de los intervalos de confianza y el coeficiente de variación, existen otros criterios importantes para validar las estimaciones estadísticas. Uno de ellos es el efecto de diseño, que mide la correlación entre la variable de interés y la distribución de los hogares en las unidades primarias de muestreo (UPM), influido por el tamaño de UPM y la correlación intraclase. También está el tamaño de muestra, que afecta la precisión de las estimaciones, sugiriendo que muestras inferiores a 100 unidades pueden ser poco fiables. Asimismo, el conteo de casos no ponderados determina la fiabilidad del estimador, clasificándose como no fiable si es menor a 30. Finalmente, los grados de libertad también son esenciales, ya que su reducción en muestras pequeñas puede afectar la precisión de las estimaciones. Estos elementos complementan el análisis de IC y CV para garantizar la validez de los resultados.

### **Especificaciones del diseño estadístico**

Para evaluar la precisión y confiabilidad de los indicadores estimados, se hace preciso conocer las especificaciones del diseño estadístico de la encuesta. En el caso de la ENCOVI 2023, se utilizó un diseño probabilístico, bietápico, estratificado y por conglomerados, tomando como dominios de estudio los 22 departamentos de Guatemala. Al trabajar con programas estadísticos, es necesario declarar las variables de diseño para obtener indicadores precisos de los parámetros de interés. Las variables que representan el diseño en la base de datos son DEPTO (departamento) y UPM (unidad primaria de muestreo), además de la variable FACTOR, que debe ser incorporada para ajustar los resultados correctamente.

### **Contenido de las bases de datos**

A continuación, se proporciona una breve descripción de las variables incluidas en cada base de datos, organizadas según los diferentes módulos temáticos de la encuesta.

## Base de agregado de consumo

La base de datos **ENCOVI.2023\_Agregado.Consumo** recoge información sobre los gastos de consumo de los hogares en Guatemala. Su propósito es proporcionar una visión integral de cómo los hogares distribuyen su consumo en distintas categorías. Esta base de datos consta de 29 variables, todas de tipo numérico, organizadas en las siguientes áreas temáticas:

**Esta base de datos contiene 10,930 observaciones, considerando únicamente los hogares que reportaron consumo.**

- **Distribución del consumo:** refleja el consumo de los hogares en alimentos, vivienda, transporte y comunicaciones, salud, educación, valor de uso de los bienes durables, gastos diarios del hogar, servicios legales, personales y otros gastos.
- **Medición de la pobreza:** incluye información sobre las líneas de pobreza extrema y total, y clasifica a los hogares en pobres extremos, pobres no extremos y no pobres, permitiendo el análisis de la situación económica y social de los hogares.
- **Variables resumen:** contienen datos como el total de kilocalorías diarias de la línea de pobreza extrema por área de residencia, el índice de precios al consumidor (IPC) para deflactar a nivel regional, el índice de Engel para cada hogar, y percentiles basados en el consumo.

## Base de personas

Esta base de datos está compuesta por 625 variables que brindan información detallada sobre los integrantes de cada hogar, permitiendo realizar análisis de sus características y condiciones. A continuación, se presenta un resumen de los temas que abordan las variables de esta base de datos:

- **Características de las personas:** incluye variables que indican el sexo, edad, dificultades físicas, relación de parentesco con el jefe del hogar, estado conyugal, código de persona, nivel educativo, idioma, autoidentificación étnica, idioma hablado, entre otras.
- **Participación en grupos comunales y uso de medios de información:** datos sobre la participación de personas de 7 años en adelante, en organizaciones, asociaciones o grupos comunitarios. Además, incluye información sobre los medios de

comunicación utilizados por personas de 18 años o más para mantenerse informadas acerca de las noticias nacionales y locales.

- **Salud:** contiene variables relacionadas con el cuidado de la niñez y la lactancia materna para niños de 0 a 24 meses, inmunizaciones y enfermedades respiratorias en menores de 6 años, acceso a servicios de salud para todas las personas, fecundidad y salud materna en mujeres de 12 a 49 años.
- **Educación:** aborda aspectos como la inscripción preescolar de niños menores de 7 años, alfabetismo, matrícula escolar para personas de 7 años o más, así como gastos mensuales y anuales en educación y asistencia escolar.
- **Capacitación para el trabajo:** incluye información sobre la capacitación laboral para personas de 12 años o más.
- **Migración:** proporciona datos sobre el lugar de nacimiento y de residencia de las personas.
- **Uso del tiempo:** reúne información sobre el tiempo invertido en trabajo remunerado y no remunerado, actividades educativas, mantenimiento del hogar, cuidado de los integrantes del hogar y el tiempo dedicado a compras y pagos.
- **Fuerza de trabajo:** contiene variables relacionadas con la ocupación principal y secundaria, subocupación, y la determinación de personas ocupadas o desocupadas.
- **Otros ingresos:** incluye información sobre ingresos no laborales en periodos de 3 y 12 meses.
- **Uso de tecnología:** proporciona información sobre el acceso y uso de dispositivos tecnológicos y de Internet.
- **Variables resumen:** este conjunto incluye variables calculadas para facilitar el análisis, como ingreso, grupos etarios, población en edad de trabajar, población económicamente activa, conteo de personas ocupadas y desocupadas, población subempleada visible, población económicamente inactiva y condición de formalidad laboral.

**Las variables relacionadas con educación, migración, uso de tiempo, fuerza de trabajo y uso de tecnología contienen información de personas de 7 años o más.**

## Base de hogares

La base de datos denominada **ENCOVI.2023\_Hogares** consta de 128 variables y 10,964 observaciones, ofreciendo una visión de las condiciones de vida de los hogares. A continuación, se presenta un resumen del contenido de las variables:

- **Características de la vivienda y el hogar:** incluye información sobre el tipo de vivienda ocupada por los hogares, materiales de paredes, techos y pisos, conexiones a redes de servicios básicos, tenencia de la vivienda, mejoras realizadas y la situación habitacional. También abarca datos sobre las fuentes de energía utilizadas por el hogar, donaciones recibidas en los últimos 12 meses y el equipamiento del hogar (bienes durables).
- **Emigración:** contiene datos relacionados con la emigración de algún miembro del hogar en los últimos 5 años.
- **Seguridad alimentaria:** reúne información sobre las experiencias de seguridad alimentaria que han vivido los hogares.
- **Seguridad ciudadana:** datos sobre la percepción de seguridad de los integrantes del hogar y las acciones tomadas ante situaciones de inseguridad.
- **Variable resumen:** esta base contiene una variable que resume la cantidad de personas residentes habituales de cada hogar.

## Base de fuentes

Esta base de datos recopila información sobre las fuentes de energía que utilizan los hogares guatemaltecos para actividades diarias como cocinar, iluminar y otras necesidades.

- **Acceso a fuentes de energía:** identificación de las fuentes de energía utilizadas por los hogares para actividades esenciales como cocinar, iluminar o calentar agua.
- **Consumo energético:** información sobre la cantidad de energía consumida por los hogares y los gastos asociados.
- **Clasificación socioeconómica:** la clasificación de los hogares según la línea de pobreza, lo que permite realizar análisis comparativos sobre el acceso a servicios energéticos entre grupos de diferentes niveles económicos.

La base contiene un total de 19 variables, todas son de tipo numérico, excepto la variable DESC\_FUENTES, que es de tipo cadena y proporciona la descripción de las fuentes de energía utilizadas por los hogares.

## Base de donaciones

Esta base de datos recopila información sobre las donaciones recibidas por los hogares, incluyendo detalles sobre el tipo de donación, como alimentos, ropa, medicinas y materiales de construcción.

- **Recepción de donaciones:** identificación de los hogares que han recibido donaciones y el tipo de bienes donados (alimentos, ropa, medicinas, etc.).
- **Valor de las donaciones:** información sobre el valor económico de las donaciones recibidas.
- **Clasificación socioeconómica:** la clasificación de los hogares según pobreza, lo que permite realizar análisis sobre la distribución de donaciones entre distintos niveles económicos.

**La encuesta identifica cuatro entidades principales que pueden proporcionar donaciones a los hogares:**

- **ONG's o grupos religiosos**
- **Instituciones internacionales**
- **Instituciones privadas**
- **Familiares, amigos y/o vecinos**

La base contiene un total de 18 variables, todas son de tipo numérico, excepto la variable DESC\_DONACIONES, que es de tipo cadena y proporciona la descripción de las donaciones recibidas por los hogares.

## Bases de emigración

Las bases de datos de emigración provienen del capítulo 1, sección I de la Encuesta Nacional de Condiciones de Vida, estas proporcionan información sobre los movimientos migratorios a nivel de hogar. Estas se dividen en dos conjuntos de datos “Emigración.1 y Emigración.2”, ambas detallan información sobre la emigración en el hogar.

La base de datos **Emigración.1** contiene un total de 10 variables, todas de tipo numérico. Recopila información sobre la emigración desde los hogares durante los últimos cinco años, registra si algún miembro del hogar se ha mudado al extranjero y la cantidad de personas que han emigrado.

La información sobre las personas que se han mudado al extranjero, incluyendo datos demográficos y de localización se encuentra en la base **Emigración.2**, esta contiene un total de 13 variables, todas de tipo numérico.

### Base de equipamiento

La base de datos equipamiento recopila información sobre los bienes durables y artículos que poseen los hogares, incluyendo artículos de cocina, artículos personales y de esparcimiento, otros artículos del hogar y vehículos, la base contiene un total de 17 variables, todas son de tipo numérico, excepto la variable "DESC\_EQUIPA", que es de tipo cadena y describe los artículos consultados en el hogar. Los temas contenidos en esta base son:

- **Propiedad de artículos:** registra si el hogar posee o cuenta con distintos artículos, tales como artículos de cocina, personales, de esparcimiento, otros artículos del hogar, y vehículos.
- **Cantidad y antigüedad:** proporciona información sobre la cantidad de artículos que posee el hogar y los años que tiene cada uno, en el caso que fuera más de un artículo del mismo tipo, se captura la información del más nuevo.
- **Condición y valor económico:** brinda información sobre si el artículo o bien adquirido era nuevo o usado y permite conocer el valor estimado de cada artículo en caso de ser vendido el día de la encuesta.

### Base de violencia

Esta base de datos está compuesta por 23 variables y aborda la seguridad personal y la protección de los bienes de los hogares, su objetivo principal es analizar el impacto de los delitos, tanto violentos como no violentos, en la calidad de vida de los hogares y sus comunidades, además de estudiar las medidas tomadas por las víctimas, como la denuncia ante las autoridades.

- **Seguridad personal y de bienes:** refleja los diferentes tipos de violencia que afectan al hogar, abarcando desde robos y asaltos hasta agresiones de mayor gravedad.

- **Acceso a la justicia:** la base recoge información sobre las denuncias que se presentan y las razones por las que, en algunos casos, no se realiza una denuncia, así como la cantidad de personas agredidas y ubicación de estas.

## Base de programas

La base de datos está compuesta por 28 variables de tipo numérico, excepto DESC\_BENEFICIO, que es una variable de tipo cadena que describe el beneficio social recibido. Los temas principales son:

- **Acceso a beneficios sociales:** permite analizar cuántos hogares reciben distintos tipos de beneficios, así como la frecuencia de estos.
- **Costo y valor de los beneficios:** facilita la evaluación de cuánto estarían dispuestos a pagar los beneficiarios por estos servicios y cómo se relaciona eso con la oferta disponible.

**Se registra un máximo de 3 beneficiarios por hogar para cada programa social.**

## Bases de gastos

La base de datos denominada **Gastos.1** contiene las variables que identifican a la persona más informada sobre los gastos del hogar. En las bases de datos **Gastos.2**, **Gastos.3** y **Gastos.4** se detalla información sobre los gastos de los hogares en transporte, comunicaciones, comidas fuera de casa, artículos para el hogar y cuidado personal, correspondientes a la semana previa a la entrevista, el mes anterior a la entrevista y los últimos doce meses, respectivamente.

## Bases de alimentos

La base de datos denominada **Alimentos.1** contiene las variables que identifican a la persona más informada sobre los gastos en alimentos del hogar. En la base **Alimentos.2** se incluye el listado de productos alimenticios consumidos por el hogar, junto con su respectiva unidad de medida. La base **Alimentos.3** registra el gasto realizado en la compra de esos alimentos durante los últimos doce meses y los últimos 15 días,

**En la ENCOVI 2023 se incluyó un listado de 116 productos alimenticios.**

además de recopilar información sobre alimentos provenientes de la producción propia o adquiridos sin compra, especificando la fuente de obtención. Finalmente, la base **Alimentos.4** contiene datos sobre los hogares que compraron alimentos, bebidas y tabaco en supermercados u otros establecimientos, pero no recuerdan el desglose exacto de los gastos.

### **Bases de negocios**

Las bases de datos **negocios.1** y **negocios.2** buscan recabar información sobre los negocios no agropecuarios que existen en los hogares. Estas bases permiten analizar aspectos relacionados con el número de negocios, el personal empleado, los ingresos, los gastos, y las actividades económicas que desarrollan estos negocios.

**Negocios.1** contiene información básica sobre si existe algún comercio y la cantidad de negocios o fábricas no agropecuarias en el hogar.

**Negocios.2** aporta información más detallada sobre los negocios, incluyendo su actividad económica, el tiempo de existencia, el propietario, los costos operativos, los ingresos, y el personal empleado (tanto remunerado como no remunerado).

### **Bases de tierra**

Las bases de datos relacionadas con la tierra proporcionan información sobre el uso, la tenencia y los derechos de propiedad de parcelas o terrenos. La base **Tierra.1** contiene la variable relacionada con la identificación de las parcelas o terrenos que los hogares tienen para su uso, ya sea con derechos de propiedad o de uso. La base **Tierra.2** recopila información sobre las parcelas o terrenos, incluyendo su nombre identificativo, extensión en manzanas cuadradas, régimen de tenencia y uso principal. Además, se detalla si la parcela cuenta con documentos legales de propiedad o arrendamiento emitidos por el Registro General de la Propiedad o Catastro, y si está a nombre del titular. También se pregunta sobre los derechos de venta y posesión legal, así como la posibilidad de que los propietarios puedan perder involuntariamente los derechos sobre la parcela en los próximos cinco años.

### **Base de unidad**

La base de datos contiene información sobre si algún miembro del hogar ha tenido tierras en los últimos 12 meses para actividades agrícolas, ganaderas o forestales, como sembrar,

cosechar, criar animales o talar árboles. Además, se especifica el número de productores agropecuarios dentro del hogar.

### **Base de producción**

En esta base de datos se detalla la información sobre las actividades agropecuarias relacionadas con el arrendamiento y venta de tierras, los ingresos generados por estas transacciones y los cultivos realizados por el hogar. Además, se especifica el tipo de cultivo sembrado, la superficie cosechada, las cantidades producidas, vendidas y reservadas para el hogar o los animales junto con su valor estimado. También se incluyen datos sobre los gastos relacionados con la producción como semillas, fertilizantes, mano de obra y maquinaria, así como las ganancias obtenidas por los cultivos. Con estos datos se pueden realizar análisis de las actividades agropecuarias y su rentabilidad en los últimos 12 meses.

### **Base de pecuarios**

La información contenida en las variables de esta base de datos trata sobre la cría de animales en el hogar, su valor de mercado y la cantidad vendida en los últimos 12 meses, junto con el valor total de esas ventas. Se tiene los gastos relacionados con la actividad, ganadera, como el pago a trabajadores, alimentos, vacunas y medicinas para animales y mantenimiento de cercas comederos y otros.

### **Base de productos**

En esta base de datos se detalla la información sobre la elaboración o recolección de productos y subproductos pecuarios en los últimos 12 meses. Incluye la cantidad mensual producida, vendida y destinada al consumo del hogar, junto con su valor estimado. Se tiene la información sobre los gastos mensuales promedio en la elaboración de subproductos y los totales de las ganancias obtenidas.

### **Base de instrumentos**

Esta base de datos contiene un total de 13 variables, la mayoría numéricas, salvo la variable "DESC\_INSTRUMENTO", que es de tipo cadena y describe el tipo de instrumento. Abarca tres aspectos fundamentales:

- **Tipo de instrumentos agropecuarios:** se identifican distintos tipos de herramientas esenciales para las labores agrícolas y pesqueras.
- **Cantidad de instrumentos:** el número de herramientas presentes en el hogar.
- **Valor estimado:** el precio aproximado al que los hogares venderían estos instrumentos en el mercado.

### Base de equipo

Esta base de datos incluye información sobre la maquinaria y equipo agropecuario que posee el hogar, se tienen datos sobre el tipo de maquinaria o equipo disponible, la cantidad actual de estos y su valor estimado de mercado en caso de venta.

### Base de instalaciones

En esta base se tiene información sobre las instalaciones agropecuarias que el hogar posee, detalla el tipo de instalaciones, la cantidad actual de cada una y una estimación del costo de construcción de dichas instalaciones.

### Base de asistencia

La base de datos de asistencia técnica se utiliza para estudiar el acceso a diferentes tipos de apoyo técnico brindados por diversas instituciones, permitiendo evaluar la cobertura y el impacto de este tipo de asistencia en la población.

La base de datos está compuesta por 14 variables, de tipo numérico excepto DESC\_ASISTENCIA, que es una variable de tipo cadena que describe el tipo de asistencia técnica recibida. Se registra si el hogar recibió asistencia técnica en los últimos 12 meses y, en caso afirmativo, se especifica si se pagó por la misma y cuál fue el monto.

- **Acceso a asistencia técnica:** permite analizar cuántos hogares han recibido asistencia técnica de diferentes fuentes, como el MAGA, cooperativas, empresas privadas, ONGs, instituciones internacionales, iglesias, personas individuales u otras instituciones gubernamentales.

- **Costo de la asistencia:** facilita la evaluación de cuánto pagaron los hogares por la asistencia técnica recibida.

## Base de forestal

La base de datos de producción forestal se utiliza para recopilar información sobre la tala y corte de árboles, así como el uso y comercialización de productos forestales en los hogares. Está compuesta por 18 variables, de tipo numérico excepto P16J02A (clase de árbol cortado o talado), que es de tipo cadena. Los temas contenidos en esta base son:

- **Tala y corte de árboles:** permite analizar cuántos y qué tipo de árboles se talaron, así como la finalidad ya sea de uso personal o venta.
- **Comercialización de productos forestales:** proporciona información sobre la cantidad de árboles vendidos, el precio de venta y las ganancias obtenidas por los hogares.
- **Costos de tala:** registra los gastos en el proceso de tala y corte, incluyendo jornales y transporte, así como el valor económico de los árboles utilizados en el hogar.

## Bases de préstamos

La base de datos **Préstamos.1** contiene información sobre los préstamos solicitados y otorgados a los miembros del hogar en los últimos 12 meses. Incluye las razones por las cuales los préstamos solicitados no fueron otorgados, si los préstamos fueron autorizados, y los motivos por los que pudieron haber sido rechazados. También se registran los préstamos vigentes o con saldo, así como el número total de préstamos que los miembros del hogar recibieron, pagaron o están pagando.

La base denominada **Préstamos.2** se centra en los detalles específicos de los préstamos recibidos. Incluye información sobre la entidad o persona que otorgó el préstamo, la forma en que se entregó, el monto recibido, y las garantías ofrecidas. Además, se registran los documentos firmados, el plazo para pagar el préstamo, la tasa de interés cobrada, los gastos administrativos asociados, y si el prestatario habría preferido un préstamo de mayor monto. También se detalla el uso principal del préstamo, la forma de pago, el número de pagos realizados o pendientes, la frecuencia de los pagos y el monto de cada cuota.

## Bases de crédito

La base de datos nombrada **Credito.1** contiene la información sobre las compras de alimentos, electrodomésticos y otros bienes realizadas al crédito por los miembros del hogar en los últimos 12 meses. Se tiene registro de los montos y cuántas compras de este tipo realizaron o están pagando en la actualidad.

La información del lugar donde se realizaron las compras, el destino de los bienes adquiridos, el enganche pagado, el monto total a pagar y las formas de pago se encuentran en la base **Credito.2**. También se detalla el costo si los bienes se hubieran pagado de contado, las garantías ofrecidas, los documentos firmados, el plazo para pagar el crédito, la tasa de interés cobrada, y si se aplicaron comisiones o gastos administrativos adicionales. Además, se incluye información sobre el número de pagos realizados o por realizar, la frecuencia de los mismos, y el monto total pagado o por pagar.

## Compatibilidad y uso en otros programas

Las bases de datos de SPSS son compatibles con una variedad de programas permitiendo al usuario usar el programa de su preferencia para realizar análisis.

- **Stata:** un programa estadístico avanzado que puede leer archivos CSV exportados desde SPSS, permitiendo análisis más especializados.
- **R Studio:** este software permite leer archivos CSV exportados de SPSS pueden ser leídos en R utilizando comandos como **read.csv** o **read\_sav**.
- **Python:** a través de bibliotecas como pandas, las bases de datos exportadas desde SPSS pueden ser manipuladas y analizadas en este lenguaje de programación.

## Ejemplo del uso de las bases de datos

Con el objetivo de orientar al usuario de mejor manera, a continuación, se muestra un ejemplo práctico de cómo se puede utilizar una base de datos en SPSS y en R Studio para generar resultados.

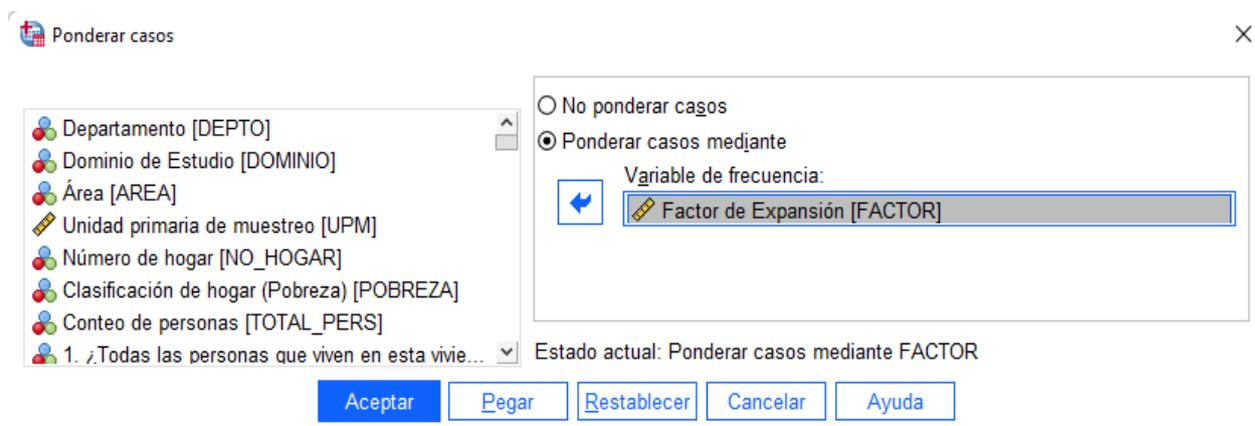
### Ejemplo usando SPSS

Para el ejemplo con SPSS se utilizará la base de hogares para realizar un análisis sobre el material predominante del techo de las viviendas donde residen los hogares guatemaltecos.

1. **Carga de la base de datos:** abrir la base de datos de hogares en SPSS y asegurarse de que la variable que describe el material del techo esté correctamente codificada como una variable categórica, para este ejemplo se utilizará la variable **P01A03**.

Nombre	Tipo	Anchura	Decimales	Etiqueta
DEPTO	Numérico	4	0	Departamento
DOMINIO	Numérico	11	0	Dominio de Estudio
AREA	Numérico	11	0	Área
UPM	Numérico	7	0	Unidad primaria de muestreo
NO_HOGAR	Numérico	8	0	Número de hogar
FACTOR	Numérico	5	0	Factor de Expansión
POBREZA	Numérico	8	2	Clasificación de hogar (Pobreza)
TOTAL_PERS	Numérico	11	0	Conteo de personas
PPB01	Numérico	3	0	1. ¿Todas las personas que viven en esta vivienda comparten un mismo gasto de servicios básic...
PPB02	Numérico	4	0	2. ¿Cuántos hogares hay en esta vivienda?
PPB03	Numérico	4	0	3. Del total de hogares que hay en esta vivienda. ¿este es el hogar número?
P01A01	Numérico	4	0	1. ¿El tipo de vivienda que ocupa el hogar es?
P01A02	Numérico	4	0	2. ¿Cuál es el material predominante en las paredes exteriores?
P01A03	Numérico	4	0	3. ¿Cuál es el material predominante en el techo?

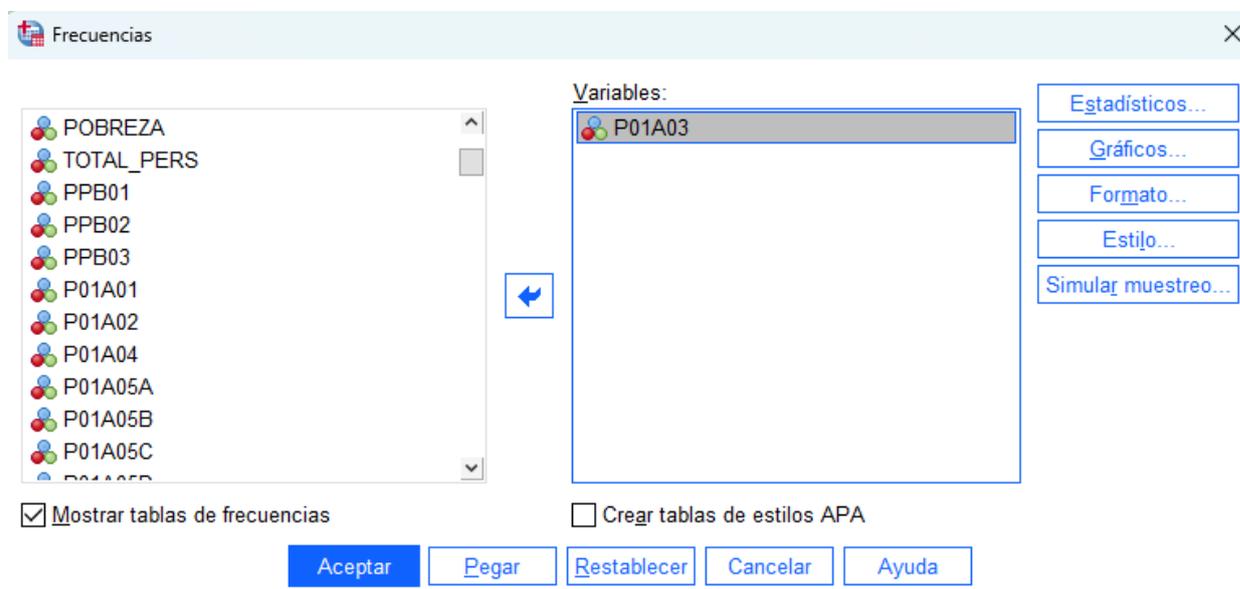
2. **Ponderación de la base de datos:** antes de realizar cualquier análisis, aplicar la ponderación seleccionando la variable llamada **FACTOR**, la cual ajustará el análisis para reflejar adecuadamente la estructura de la población, siguiendo estos pasos: **Datos > Ponderar Casos > Factor > Aceptar**; o también puede realizarse usando la sintaxis **"WEIGHT BY FACTOR"**.



3. **Análisis descriptivo del material del techo:** una vez aplicada la ponderación, se genera un análisis descriptivo para observar la distribución de frecuencias de los diferentes materiales predominantes del techo, esto puede realizarse mediante una tabla de frecuencias. Para generar esta tabla se pueden seguir los siguientes pasos: **[Analizar > Estadísticos Descriptivos > Frecuencias > P01A03](#)** o también se puede utilizar el siguiente código de sintaxis:

***FRECUENCIES VARIABLES=P01A03***

***/ORDER=ANALYSIS.***



La salida que obtendrá es la siguiente:

3. ¿Cuál es el material predominante en el techo?					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Concreto	807674	19,9	19,9	19,9
	Lámina metálica	3025837	74,4	74,4	94,2
	Asbesto cemento	48024	1,2	1,2	95,4
	Teja	137722	3,4	3,4	98,8
	Paja, palma o similar	47170	1,2	1,2	99,9
	Otro, cual?	2429	,1	,1	100,0
	Total	4068856	100,0	100,0	

4. **Desagregación del análisis:** si se desea analizar cómo varía el material del techo según variables como departamento o área, se pueden realizar cruces de variables de la siguiente manera **Analizar > Tablas > Tablas Personalizadas > Área o Departamento** o puede utilizar la siguiente sintaxis:

**CTABLES**

**/VLABELS VARIABLES=AREA P01A03 DISPLAY=BOTH**

**/TABLE AREA [COUNT F40.0] BY P01A03**

**/CATEGORIES VARIABLES=AREA P01A03 ORDER=A KEY=VALUE EMPTY=INCLUD**

**/CRITERIA CILEVEL=95.**

## Resultado

		P01A03 3. ¿Cuál es el material predominante en el techo?					
		1 Concreto Recuento	2 Lámina metálica Recuento	3 Asbesto cemento Recuento	4 Teja Recuento	5 Paja, palma o similar Recuento	98 Otro, cual? Recuento
AREA Área	1 Urbana	689417	1531981	39598	53964	3604	562
	2 Rural	118257	1493856	8426	83758	43566	1867

**Nota.** Para realizar un cruce con otra variable únicamente se sustituye la palabra "AREA" por la variable seleccionada.

**5. Exportación de la base de datos:** SPSS ofrece la posibilidad de exportar la base de datos para su uso en otros programas de análisis o procesamiento de datos.

Para exportar a Excel ir a **Archivo > Exportar**, seleccionar el formato **"Excel (.xls o .xlsx)"** y elegir la ubicación para guardar el archivo.

Para Exportar a CSV, un formato más universal, seleccionar formato **CSV (Valores Separados por Comas)**.

### Ejemplo utilizando R Studio

En este ejemplo se ilustra cómo utilizar R Studio para realizar cálculos y análisis de datos a partir de una base de datos en formato SPSS. Se abordan los pasos necesarios para la carga de librerías, importación de datos, construcción de códigos específicos y exportación de resultados. Para el desarrollo de este ejemplo se utilizará la base de agregado de consumo.

- 1. Carga de librerías necesarias:** para importar una base de datos de SPSS a R, es necesario utilizar la librería "haven". Para un análisis completo de datos, se puede emplear la librería "tidyverse". El siguiente código permite cargar las librerías:

```
library(tidyverse)  
library(haven)  
library(writexl)
```

Si las librerías aún no se encuentran descargadas en el entorno de R, es necesario descargarlas usando el siguiente código:

```
install.packages(c("tidyverse", "haven", "writexl"))
```

- 2. Carga de la base de datos:** para cargar la base de datos existen diversas maneras puede ser mediante código o en la interfaz, sin embargo, para este ejemplo se utilizará el siguiente código de la librería "haven":

```
Consumo <- read_sav("C:/Tu/ubicación/ENCOVI.2023_Agregado.Consumo.SAV")
```

**3. Elaboración del código para cálculo de indicadores:** al tener la base de datos cargada se procede a la construcción del código que permitirá mostrar los cálculos deseados. Para efectos de este ejemplo, se estará calculando la proporción de la población que vive por debajo del umbral nacional de pobreza por área (urbana, rural). El código que se empleará es el siguiente:

```
# Calcular el total de personas por área (urbana y rural)
total_area <- Consumo %>%
  group_by(AREA) %>%
  summarise(total = sum(Factor3), .groups = 'drop')
```

Se crea un resumen para cada área calculando la suma ponderada de la variable *Factor3*, que representa el factor de expansión o el peso del registro en la muestra. El valor resultante se almacena en una nueva columna llamada *total*.

```
# Calcular la proporción de pobreza por área
Proporciones <- Consumo %>%
  filter(pobreza %in% c(1,2)) %>%
  group_by(AREA) %>%
  summarise(Pobreza_total = sum(Factor3), .groups = 'drop') %>%
  left_join(total_area, by = "AREA") %>%
  mutate(Proporcion = round((Pobreza_total / total) * 100, 1)) %>%
  select(AREA, Proporcion)
```

Con el código anterior, se seleccionan únicamente los registros del dataframe *Consumo* en los que la variable *pobreza* tiene un valor de 1 o 2. Esto indica que se está considerando a los hogares que se encuentran en situación de pobreza (pobres y pobres extremos). Se agrupan los datos filtrados por la variable *AREA*, que representa las áreas geográficas (en este caso, urbana y rural).

Se calcula la suma ponderada de los hogares en pobreza (*sum(Factor3)*) para cada área geográfica (*AREA*). El valor resultante se almacena en una nueva columna llamada *Pobreza total*. Se realiza una unión (*left join*) con el primer dataframe llamado *total área*, que contiene el total de hogares o personas por cada área.

Se calcula la proporción de pobreza por área dividiendo el total de hogares en situación de pobreza (*Pobreza total*) entre el total de hogares o personas (*total*) para cada área. El resultado se multiplica por 100 para obtener un valor en porcentaje y se redondea a 1 decimal. Se seleccionan las columnas *AREA* y *Proporción* para el resultado final, dejando fuera cualquier otra columna que pudiera haberse generado en los pasos anteriores.

- 4. Generación de resultado:** en esta etapa se ejecutan los cálculos y se muestran los resultados, que pueden visualizarse directamente en la consola o en un dataframe para mayor claridad, si se prefieren representaciones visuales, se pueden generar gráficos utilizando la librería “ggplot2”. El código para visualización puede utilizar el siguiente:

```
# dentro del () debe indicar en nombre del dataframe donde guardó sus resultados.
```

```
View (Proporciones)
```

- 5. Exportación del resultado:** R permite exportar datos a diversos formatos, como Excel, CSV, Stata, SPSS, entre otros. Generalmente, una gran cantidad de resultados se exportan a Excel debido a su facilidad de uso y amplia aceptación. A continuación se presenta el código para exportar un data frame a un archivo Excel

```
write_xlsx(Proporciones, "ruta de acceso donde se guardará")
```



Instituto Nacional  
de Estadística



Más información en  
***ine.gob.gt***



Encuesta Nacional de  
**Condiciones de Vida**