

Diccionario y Revisión: Base de Datos ENUSC Consolidada (2016-2024)

Joaquín Valenzuela

2025-11-19

Contents

1	1. Introducción	1
1.1	1.1 Carga de Datos	1
2	2. Variables de Identificación y Diseño	2
3	3. Variables de Victimización (Target)	3
3.1	3.1 Indicadores Agregados	3
3.2	3.2 Victimización Específica (Los 7 Delitos)	4
4	4. Variables de Denuncia	4
5	5. Variables Sociodemográficas	4
6	6. Percepción y Contexto	5
7	7. Resumen de Valores Perdidos (NAs)	5

1 1. Introducción

Este documento detalla la construcción, estructura y contenido de la base de datos maestra `base_datos_final.rds`.

El objetivo de esta base es unificar 9 años de la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC), armonizando los cambios de nombres de variables y escalas de respuesta ocurridos entre 2016 y 2024.

1.1 1.1 Carga de Datos

Cargamos la base de datos procesada para realizar la inspección.

```

# Ajusta esta ruta si es necesario
ruta_base <- "data/base_datos_final.rds"

if(file.exists(ruta_base)) {
  df <- readRDS(ruta_base)
  cat("superB ase cargada exitosamente.\n")
  cat("Dimensiones:", nrow(df), "observaciones y", ncol(df), "variables.")
} else {
  stop("error No se encuentra el archivo. Ejecuta primero el script 01_crear_base.R")
}

## superB ase cargada exitosamente.
## Dimensiones: 247814 observaciones y 33 variables.

```

2 2. Variables de Identificación y Diseño

Estas variables sitúan al hogar en el tiempo y el espacio, y proveen los ponderadores necesarios para la inferencia poblacional.

2.0.1 anio

- **Descripción:** Año de levantamiento de la encuesta.
- **Origen:** Extraído del nombre del archivo original (`enusc2016.csv`, etc.).
- **Rango:** 2016 a 2024.

```
table(df$anio)
```

```

##
## 2016 2017 2018 2019 2020 2021 2022 2023 2024
## 25227 27616 27025 24465 23656 22180 23360 49813 24472

```

2.0.2 region

- **Descripción:** Región de residencia del hogar.
- **Origen:** Armonización de las variables `enc_region` (años recientes) y `enc_region16` (años antiguos).
- **Escala:** Códigos numéricos oficiales (1-16). (13 = Región Metropolitana).

2.0.3 fact_hog

- **Descripción:** Factor de expansión del hogar. Indica a cuántos hogares representa cada observación.
- **Origen:** Variables `fact_hog`, `fact_hog_reg` o variantes según el año.
- **Nota:** Vital para obtener totales nacionales.

3 3. Variables de Victimización (Target)

Estas variables fueron **construidas** a partir de las preguntas crudas de la encuesta. Aquí se detalla la “cocina” de cada indicador.

3.1 3.1 Indicadores Agregados

3.1.1 hogar_victima_dmcs (Variable Principal)

- **Descripción:** Indica si el hogar fue víctima de **al menos uno** de los 7 Delitos de Mayor Connotación Social (DMCS) en los últimos 12 meses.
- **Lógica de Construcción:**
 1. Se identifican los 7 delitos individuales.
 2. Se suman los eventos: $n_delitos = RVI + RPS + RFV + HUR + LES + RDV + RDDV$.
 3. Si $n_delitos > 0 \rightarrow 1$ (Víctima). Si no $\rightarrow 0$.
- **Escala:** 1 = Sí, 0 = No.

```
prop.table(table(df$hogar_victima_dmcs)) * 100
```

```
##  
##      0       1  
## 79.13798 20.86202
```

3.1.2 cifra_oscura_dmcs (Variable Dependiente del Modelo)

- **Descripción:** Identifica a los hogares que, habiendo sido víctimas, **NO denunciaron** el hecho.
- **Lógica de Construcción:**
 1. Se filtra solo a quienes tienen `hogar_victima_dmcs == 1`.
 2. Se verifica si realizaron al menos una denuncia (`hogar_denuncio_dmcs`).
 3. Cálculo inverso: $1 - \text{hogar_denuncio_dmcs}$.
- **Escala:** * 1 = Es Cifra Oscura (No denunció).
 - 0 = Denunció.
 - NA = No aplica (No fue víctima).

```
# Solo para víctimas  
summary(df$cifra_oscura_dmcs)
```

```
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.    NA's  
##   0.00   0.00   1.00   0.64   1.00   1.00 196115
```

3.2 3.2 Victimización Específica (Los 7 Delitos)

Para cada delito, se aplicó una **regla de armonización binaria** para corregir el cambio de escala de la ENUSC (donde antes “No” era 2 y ahora es 0).

Regla aplicada en R: `if_else(variable_raw == 1, 1, 0)` Esto asegura que solo el “Sí” explícito cuente como 1.

1. **victima_rvi:** Robo con Violencia o Intimidación.
 2. **victima_rps:** Robo por Sorpresa.
 3. **victima_rfV:** Robo con Fuerza en la Vivienda.
 4. **victima_hur:** Hurto.
 5. **victima_les_agr:** Lesiones o Agresiones.
 6. **victima_rdv:** Robo de Vehículo Motorizado.
 7. **victima_rddv:** Robo de Accesorios de/desde Vehículo.
-

4 4. Variables de Denuncia

Estas variables indican si se denunció cada delito específico. Son la base para calcular la Cifra Oscura.

Lógica de Construcción: Son condicionales. Solo tienen valor (0 o 1) si la persona fue víctima de ese delito específico.

- `if (victima_X == 1) →` mirar variable de denuncia (1=Sí, 0=No).
- `if (victima_X == 0) → NA` (No aplica).

Lista de variables:

- `denuncio_rvi, denuncio_rps, denuncio_rfV, denuncio_hur, denuncio_les_agr, denuncio_rdv, denuncio_rddv.`
-

5 5. Variables Sociodemográficas

Estas variables provienen del módulo “Datos del Hogar” o del “Informante Kish” seleccionado.

5.0.1 sexo

- **Escala:** 1 = Hombre, 2 = Mujer.

5.0.2 edad_cat (Variable a corregir en el Modelo)

- **Descripción:** Edad del informante.
- **Nota de Procesamiento:** Contiene una mezcla de **edades exactas** (ej. 45, 60) y **códigos de tramos** (ej. 1, 2, 3) debido a cambios en el cuestionario por año.
- **Acción Requerida:** Debe ser recodificada a tramos estandarizados (ej. “15-29”, “30-49”) antes de entrar al modelo Heckman (esto se hace en el script 02).

5.0.3 educ (Nivel Educacional)

- **Origen:** Variable rph_nivel.
- **Escala:** Numérica ascendente (0=Sin estudios a 12/13=Postgrado).

5.0.4 nse (Nivel Socioeconómico)

- **Origen:** Quintiles o clasificación NSE del hogar.
 - **Nota:** Presenta un alto porcentaje de valores perdidos (NA) en algunos años.
-

6 6. Percepción y Contexto

VARIABLES subjetivas sobre la seguridad y confianza institucional. **Procesamiento:** Se aplicó la función na_codes() para convertir las respuestas “No sabe” (88) y “No responde” (99) en NA, evitando sesgos numéricos.

6.0.1 Confianza Institucional

Escala Likert invertida (generalmente 1=Mucha confianza, 4=Nada).

- conf_carab: Carabineros.
- conf_pdi: PDI.
- conf_fisc: Fiscalía (Ministerio Público). **Nota:** Alta tasa de no respuesta.

6.0.2 Entorno

- aum_del_barrio: Percepción de si la delincuencia aumentó en el barrio.
 - tiempo_res: Antigüedad residencial en la comuna.
-

7 7. Resumen de Valores Perdidos (NAs)

Revisión rápida de la calidad de los datos para el modelamiento.

```
na_summary <- sapply(df, function(x) sum(is.na(x)))
na_summary[na_summary > 0]
```

```
## cifra_oscura_dmcs      varstrat      re_victima      educ
##          196115          247814          196115         23656
##          nse      estado_civil      conf_carab      conf_pdi
##          93661           119009          97499        115466
##          conf_fisc     aum_del_barrio     tiempo_res    denuncio_rvi
##          135761            5655           69526        239662
##          denuncio_rps    denuncio_rfv    denuncio_hur  denuncio_les_agr
##          240421            238552          232405        244896
##          denuncio_rdv    denuncio_rddv
##          246137            231617
```

¿Qué hace este código?

1. **Se conecta automáticamente** a tu carpeta ‘data’ y carga el archivo ‘.rds’.
2. **Imprime estadísticas reales** de tus datos (tablas de frecuencia, promedios).
3. Genera un **PDF profesional** con títulos, negritas y explicaciones metodológicas que puedes anexar