

PROCESO DE ANONIMIZACIÓN – BASE DE DATOS DE GRILLAS DE 1 KM2 A NIVEL NACIONAL

Estructura de la base de datos

La base de datos está compuesta por 1.147.103 grillas de 1 km2 a nivel nacional y 8 variables. Las dos primeras variables son identificadores únicos para cada grilla y las 6 variables restantes describen cuántas viviendas, cuántos hogares y cuántas personas hay en la grilla.

Análisis de riesgos de identificación de las grillas

Para el proceso de anonimización de la base de datos se propone un posible escenario de riesgo de identificación, el cual considera que todas las grillas que contengan menos de 3 viviendas son riesgosas.

Se presenta la distribución de frecuencias de la variable Total de Viviendas (*viviendas*) dentro de cada grilla:

Total viviendas por grilla	Número de grillas	Porcentaje	Porcentaje acumulado
0	232.269	20,2%	20,2%
4	6.470	0,6%	20,8%
5	7.270	0,6%	21,4%
6 o más	108.712	9,5%	30,9%
Sin información de viviendas	792.382	69,1%	100,0%
Total grillas	1.147.103	100,0%	

En este caso se observa que, dada la recategorización del número de viviendas para el 20.2% de las grillas (grillas con 1, 2 o 3 viviendas) el riesgo de identificación en la base de datos se minimiza y está controlado.