**🧾 Variables comunes (presentes en ambos datasets)**

Estas variables están **tanto en train\_hogares como en test\_hogares**, y todas están documentadas en el PDF:

| **Variable** | **Descripción (según DANE)** |
| --- | --- |
| **id** | Identificador único del hogar (combinación de *Directorio* y *Secuencia\_p*). |
| **Clase** | Tipo de área: 1 = Cabecera (urbano), 2 = Resto (rural). |
| **Dominio** | Dominio geográfico de inferencia (área metropolitana, otras cabeceras, resto). |
| **P5000** | Número total de habitaciones del hogar (excepto cocina, baño y garaje). |
| **P5010** | Número de habitaciones utilizadas para dormir. |
| **P5090** | Tipo de tenencia de la vivienda (propia, arrendada, usufructo, etc.). |
| **P5100** | Cuota mensual por compra o amortización de vivienda (COP). |
| **P5130** | Valor estimado del arriendo imputado (si fuera en arriendo). |
| **P5140** | Valor mensual efectivamente pagado por arriendo (COP). |
| **Nper** | Número de personas del hogar. |
| **Npersug** | Número de personas en la unidad de gasto (*unidad de gasto compartida*). |
| **Li** | Línea de indigencia (ingreso mínimo para cubrir alimentación). |
| **Lp** | Línea de pobreza (ingreso mínimo para cubrir alimentación y otros bienes básicos). |
| **Fex\_c** | Factor de expansión anualizado (ponderador nacional). |
| **Depto** | Código del departamento de residencia. |
| **Fex\_dpto** | Factor de expansión departamental (ponderador departamental). |

✅ Estas 16 variables están documentadas en el archivo DANE y aparecen en ambos datasets.

**📘 Variables solo en train\_hogares (no aparecen en test\_hogares)**

Estas variables **no están** en el PDF original del DANE, sino que fueron **creadas o derivadas** durante el procesamiento para entrenamiento o análisis de pobreza:

| **Variable** | **Interpretación (derivada o adicional)** |
| --- | --- |
| **Ingtotug** | *Ingreso total de la unidad de gasto* (suma de todos los ingresos del hogar que comparten gastos). |
| **Ingtotugarr** | *Ingreso total de la unidad de gasto ajustado por arriendo imputado* — corrige el ingreso incorporando el valor del arriendo estimado para propietarios. |
| **Ingpcug** | *Ingreso per cápita de la unidad de gasto* — se calcula como Ingtotugarr / Npersug. |
| **Pobre** | Indicador binario (dummy) de pobreza monetaria: 1 si Ingpcug < Lp, 0 en caso contrario. |
| **Indigente** | Indicador binario de pobreza extrema: 1 si Ingpcug < Li, 0 en caso contrario. |
| **Npobres** | Número de personas pobres dentro del hogar (posiblemente Pobre \* Npersug). |
| **Nindigentes** | Número de personas en pobreza extrema dentro del hogar (posiblemente Indigente \* Npersug). |

🔹 Estas variables no están descritas en el PDF porque **son construcciones analíticas** derivadas del cálculo de pobreza monetaria, no variables recolectadas en la GEIH.

**📊 Resumen**

| **Tipo de variable** | **En train\_hogares** | **En test\_hogares** | **Documentada en DANE** |
| --- | --- | --- | --- |
| Variables básicas GEIH / DANE | ✅ | ✅ | ✅ |
| Líneas de pobreza (Li, Lp) | ✅ | ✅ | ✅ |
| Factores de expansión (Fex\_c, Fex\_dpto) | ✅ | ✅ | ✅ |
| Variables derivadas de ingreso | ✅ | ❌ | ❌ |
| Indicadores de pobreza (Pobre, Indigente, etc.) | ✅ | ❌ | ❌ |

En resumen:

* 🟩 **test\_hogares** contiene solo las **variables originales del DANE** (estructurales).
* 🟦 **train\_hogares** incluye esas **+ las variables derivadas** de ingreso y pobreza, que se usan para modelar, predecir o clasificar hogares según su condición económica.