**🧾 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.
* **🧾 Identificación y contexto**

| **Variable** | **Descripción** |
| --- | --- |
| **id** | Identificador único del hogar o persona. Permite empalmar con la base de hogares. |
| **Orden** | Número de orden del individuo dentro del hogar. |
| **Clase** | Tipo de área: 1 = Cabecera (urbano), 2 = Resto (rural). |
| **Dominio** | Dominio geográfico (región, departamento o ciudad según el diseño muestral). |
| **Depto** | Código del departamento. |
| **Fex\_c** | Factor de expansión individual (ponderador muestral). |
| **Fex\_dpto** | Factor de expansión departamental. |

* **👤 Características sociodemográficas**

| **Variable** | **Descripción** |
| --- | --- |
| **P6020** | Sexo del individuo: 1 = Hombre, 2 = Mujer. |
| **P6040** | Edad en años cumplidos. |
| **P6050** | Parentesco con el jefe(a) del hogar (1 = Jefe, 2 = Cónyuge, 3 = Hijo/a, etc.). |
| **P6090** | Afiliación a salud: 1 = Sí, 2 = No. |
| **P6100** | Régimen de salud: 1 = Contributivo, 2 = Especial, 3 = Subsidiado, 4 = No sabe. |
| **P6210** | Nivel educativo más alto alcanzado. |
| **P6210s1** | Último grado aprobado en el nivel educativo. |
| **P6240** | Actividad principal la semana pasada (trabajó, buscó trabajo, estudió, etc.). |

* **💼 Empleo e ingresos laborales**

| **Variable** | **Descripción** |
| --- | --- |
| **Oficio** | Ocupación principal o tipo de trabajo. |
| **P6426** | Tiempo (en meses) que lleva trabajando de forma continua en el empleo actual. |
| **P6430** | Posición ocupacional: empleado, obrero, independiente, patrón, etc. |
| **P6510** | Ingreso mensual laboral neto del empleo principal. |
| **P6545** | Ingreso mensual del segundo trabajo o actividad secundaria. |
| **P6580** | Ingreso por horas extras, bonificaciones u otros pagos ocasionales. |

* **💰 Ingresos no laborales y transferencias**

| **Variable** | **Descripción** |
| --- | --- |
| **P6585s1–P6585s4** | Componentes del ingreso no laboral: arriendos, intereses, pensiones, transferencias, remesas, ayudas. |
| **P6590** | Ingreso por jubilaciones o pensiones. |
| **P6600–P6630s6** | Otros ingresos ocasionales o transferencias (bonificaciones, subsidios, herencias, etc.). |
| **P6800**, **P6870**, **P6920** | Ayudas o ingresos en especie recibidos por el hogar. |
| **P7040–P7160** | Transferencias recibidas (subsidios, remesas, becas, ayudas sociales). |
| **P7310**, **P7350**, **P7422**, **P7472**, **P7495** | Ingresos no laborales diversos (premios, ventas, herencias, indemnizaciones, etc.). |
| **P7500s2**, **P7500s3**, **P7505** | Transferencias o ayudas recibidas por otras fuentes. |
| **P7510s1–P7510s7** | Transferencias o ingresos no laborales según tipo de fuente (por ejemplo, gobierno, familiares, entidades privadas). |

* **👷‍♂️ Condición laboral**

| **Variable** | **Descripción** |
| --- | --- |
| **Pet** | Población en edad de trabajar (1 = Sí, 0 = No). |
| **Oc** | Ocupado (1 = Sí, 0 = No). |
| **Des** | Desocupado (busca trabajo). |
| **Ina** | Inactivo (no trabaja ni busca trabajo). |