



Tratamiento de información para modelos estadísticos



OBSERVATORIO
INMOBILIARIO
CATASTRAL

Abril 2026

Contenido

1) Motivación

2) Descripción
general -
Modelos



3) Flujo de trabajo

4) Base de
modelación



5) Estructura
propuesta

6) Modelos en la
literatura

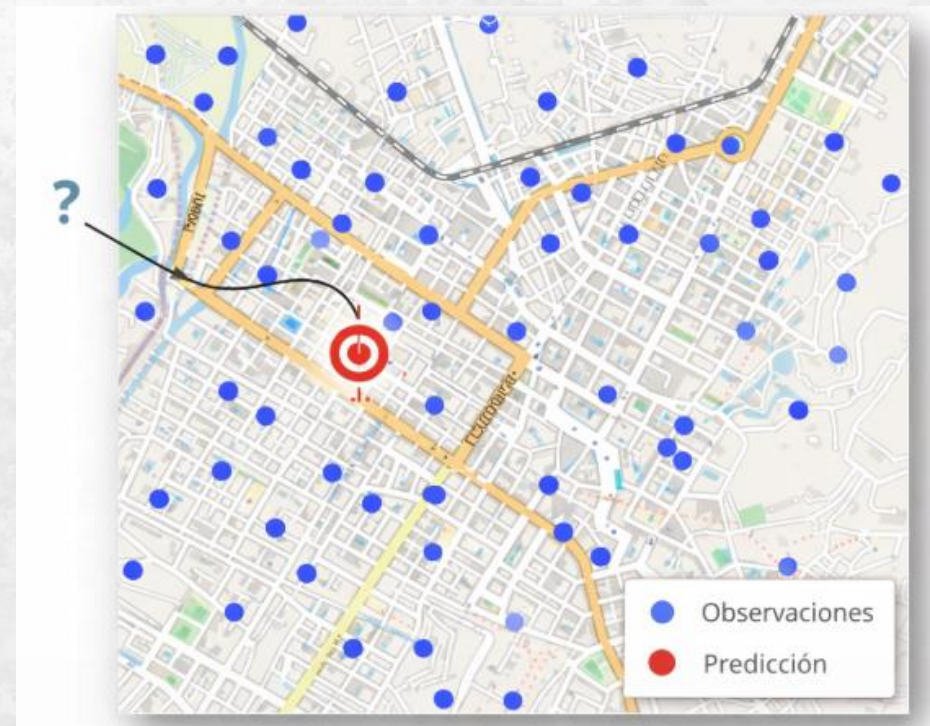
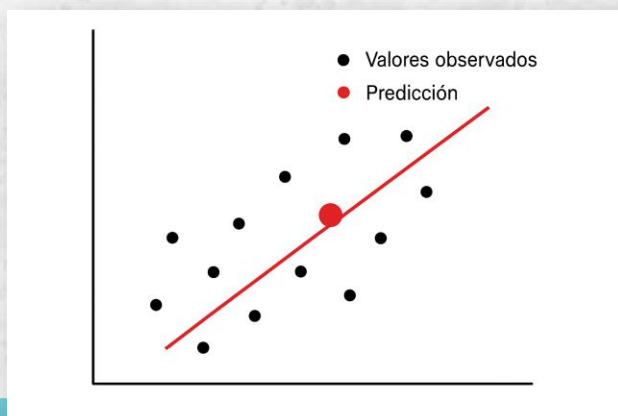


1) Problema de modelación: predicción de valores no observados

Dentro de los procesos de actualización catastral, se tienen en cuenta los componentes **económico**, jurídico y físico. Dentro de esta presentación se abordará una metodología para apoyar el primero.

- Se cuenta con observaciones georreferenciadas (valores catastrales).
- No se dispone de información en todos los puntos del territorio.
- Existe dependencia espacial: puntos cercanos tienden a tener valores similares.

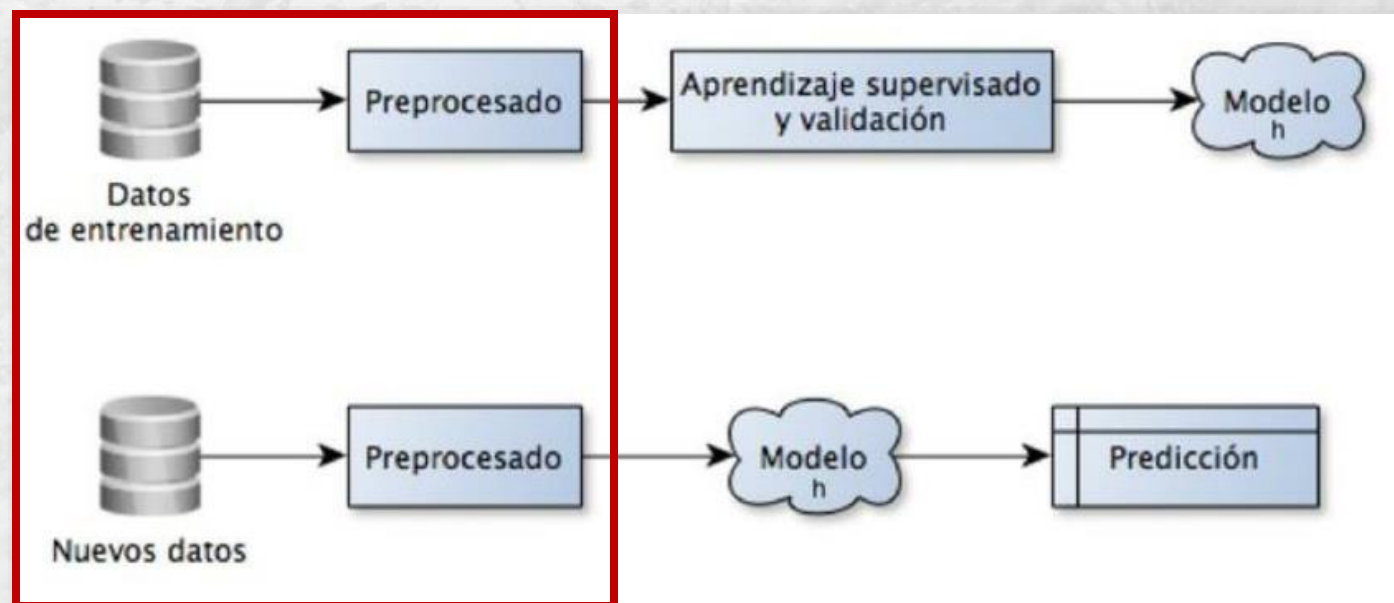
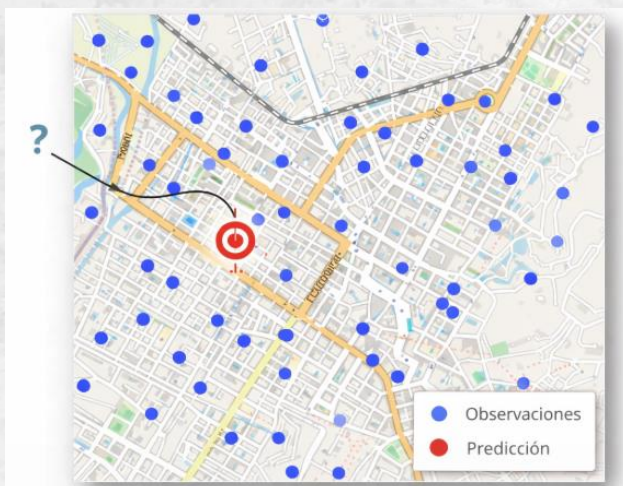
¿Cómo estimar el valor en ubicaciones sin observación a partir de datos disponibles?



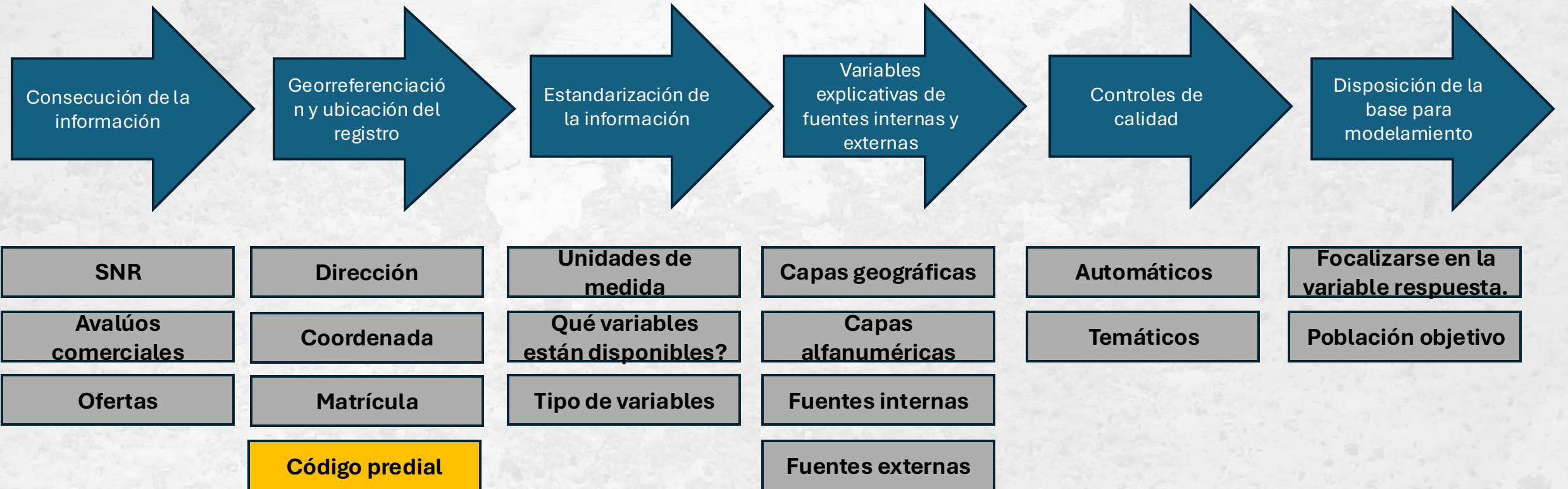
2) Descripción general - Modelos

Definición de la **variable de interés**, el **universo de análisis** y la **población objetivo**.

Ej: Valor del terreno en predios urbanos de un municipio, a partir de avalúos puntuales y considerando criterios de selección del universo de análisis (por ejemplo, excluyendo estadios, coliseos, avalúos puntuales, zonas de protección, etc.).



3) Flujo de trabajo para el tratamiento de información

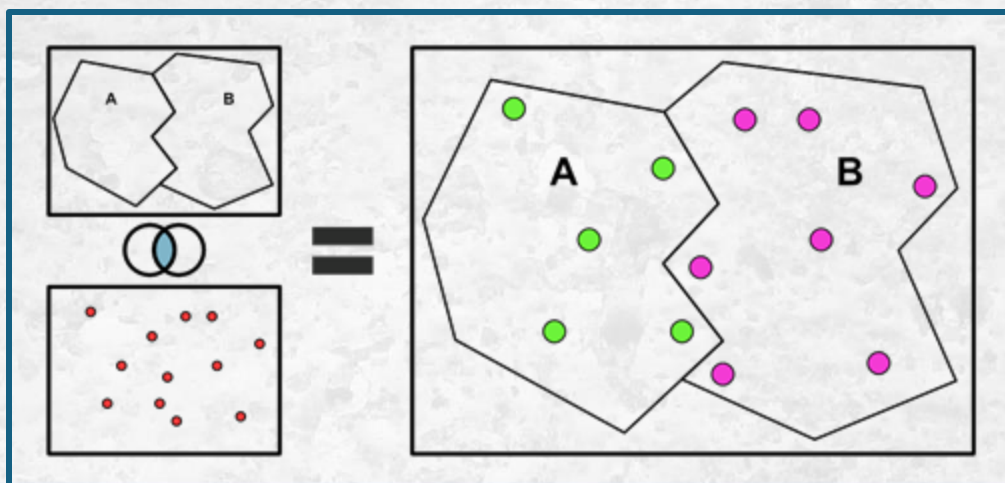


4) Base de modelación - Ejemplo

Variables geográficas

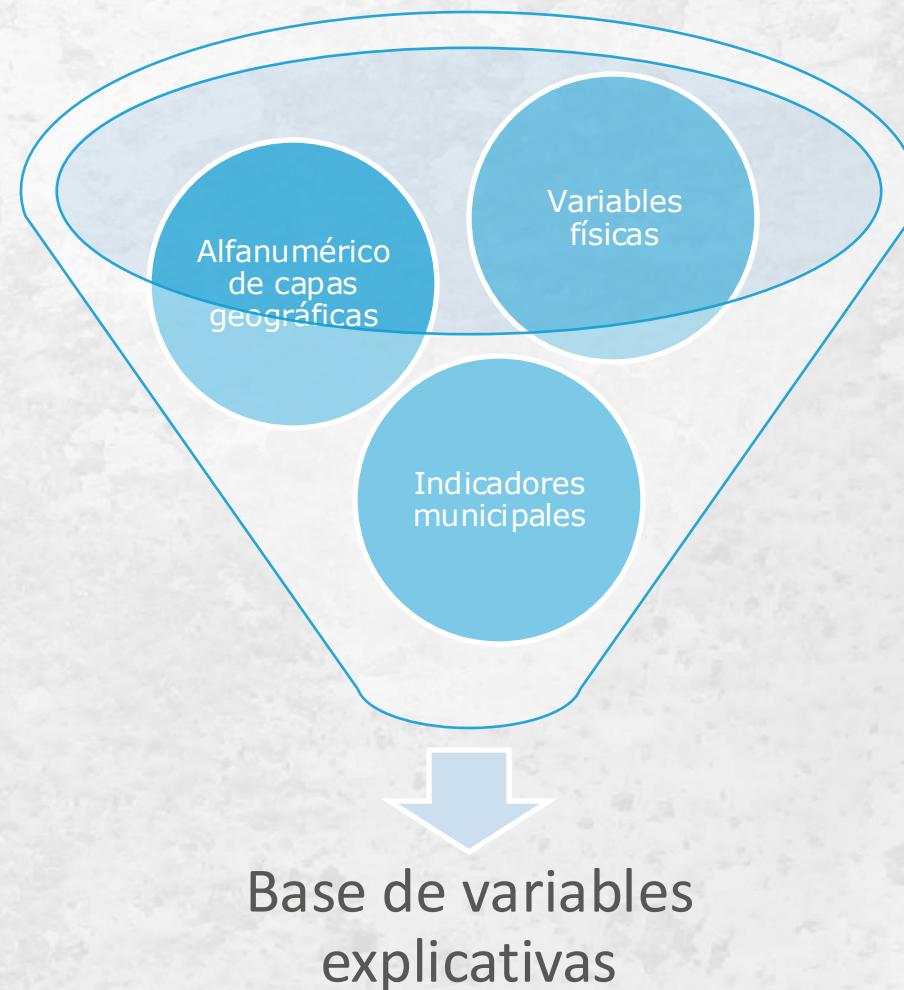
Las capas geográficas se procesan previamente y mediante un join geográfico punto-polígono, se relaciona el código predial con la información asociada de las capas.

- Zonas Homogéneas Físicas.
- Zonas Homogéneas Geoeconómicas.
- Frontera Agrícola.
- Conflicto Uso del Suelo.
- Áreas Homogéneas de Tierra.
- Capacidad Uso del Suelo.
- Uso principal del Suelo.
- Vocación uso del Suelo.
- Sistema Nacional de Áreas Protegidas.
- Parque Natural Nacionales.
- Resguardos Indígenas.
- Terreno Rural.
- Corine Land Cover.
- Clase Agrologica.



Por predio se toma la característica predominante (mayor área).

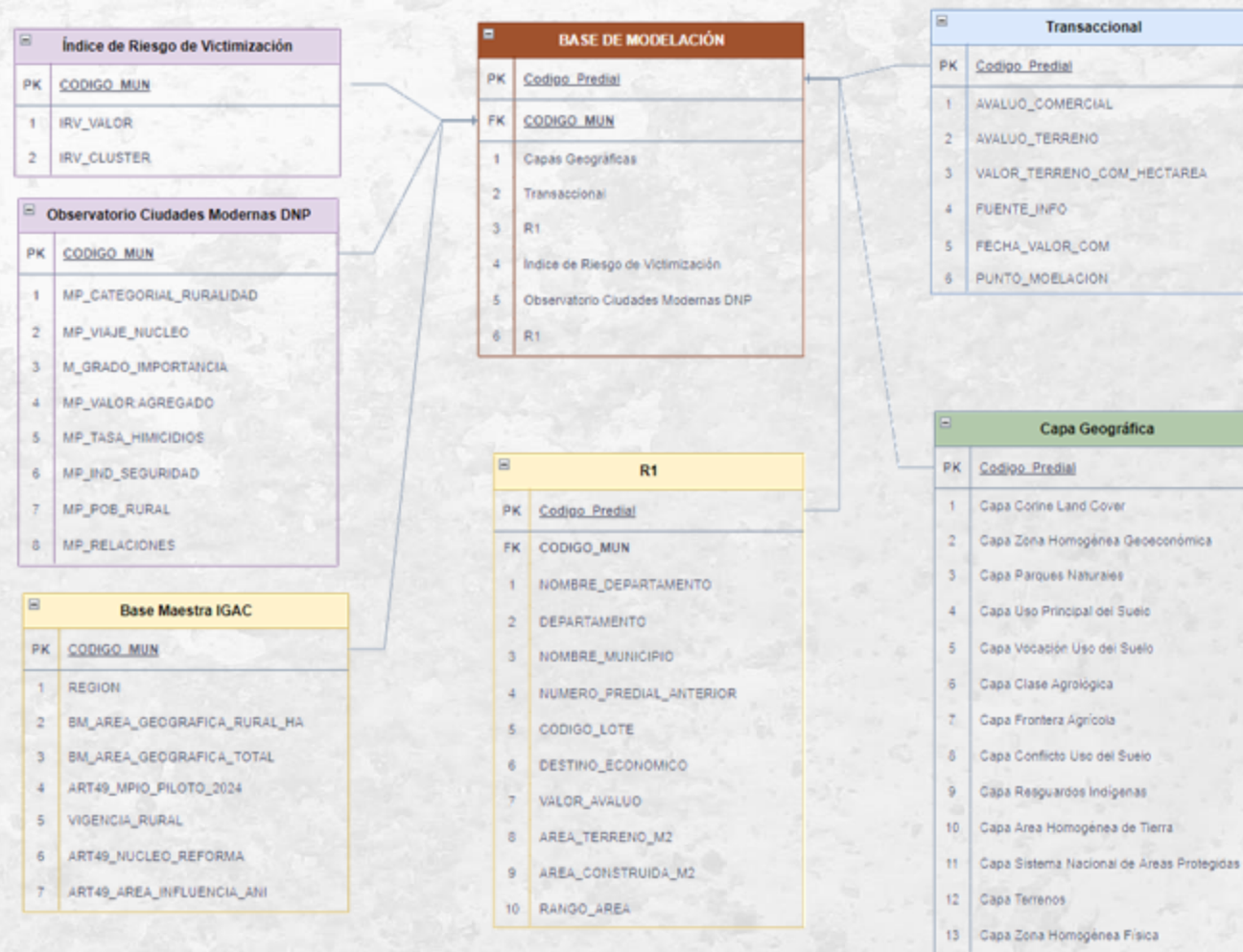
Variables exógenas



4) Base de modelación - Ejemplo

Consolidación

Finalmente, la unión de las bases se realiza teniendo en cuenta un cruce nivel predial entre el R1, la base de valores comerciales, los alfanuméricos de las capas geográficas y el cruce a nivel municipal con la información de fuentes externas y la base maestra.



5) Estructura propuesta de modelos

El proceso de modelamiento propuesto se basa en la siguiente formulación:

$$Z(\mathbf{s}) = f(\mathbf{X}(\mathbf{s}), \mathbf{s}) + \varepsilon(\mathbf{s})$$

Donde:

\mathbf{s} : es una coordenada

$Z(\mathbf{s})$: es el valor de hectárea de terreno.

$f(\mathbf{X}(\mathbf{s}), \mathbf{s})$: representa el componente de regresión

$\varepsilon(\mathbf{s})$: representa un proceso geoestadístico (espacialmente dependiente).

A diferencia de los modelos tradicionales (regresión normal), estos permiten:

- Incorporar la dependencia espacial.
- Modelar relaciones no lineales.

Proceso de modelamiento de
 $Z(\mathbf{s})$

Modelos predictivos

Validación cruzada
generalizada

Selección del mejor modelo para
el componente $f(\mathbf{X}(\mathbf{s}), \mathbf{s})$

Selección del modelo teórico de
semivariograma para $\varepsilon(\mathbf{s})$

Recolección de métricas a partir de
residuales:

$$r(\mathbf{s}) = Z(\mathbf{s}) - \hat{f}(\mathbf{X}(\mathbf{s}), \mathbf{s}) - \varepsilon(\mathbf{s})$$

6) Estructura del componente espacial

Posterior al ajuste del componente de regresión, con el modelo ganador seleccionado, se calcula:

$$\varepsilon(\mathbf{s})=Z(\mathbf{s})-f(\mathbf{X}(\mathbf{s}),\mathbf{s})$$

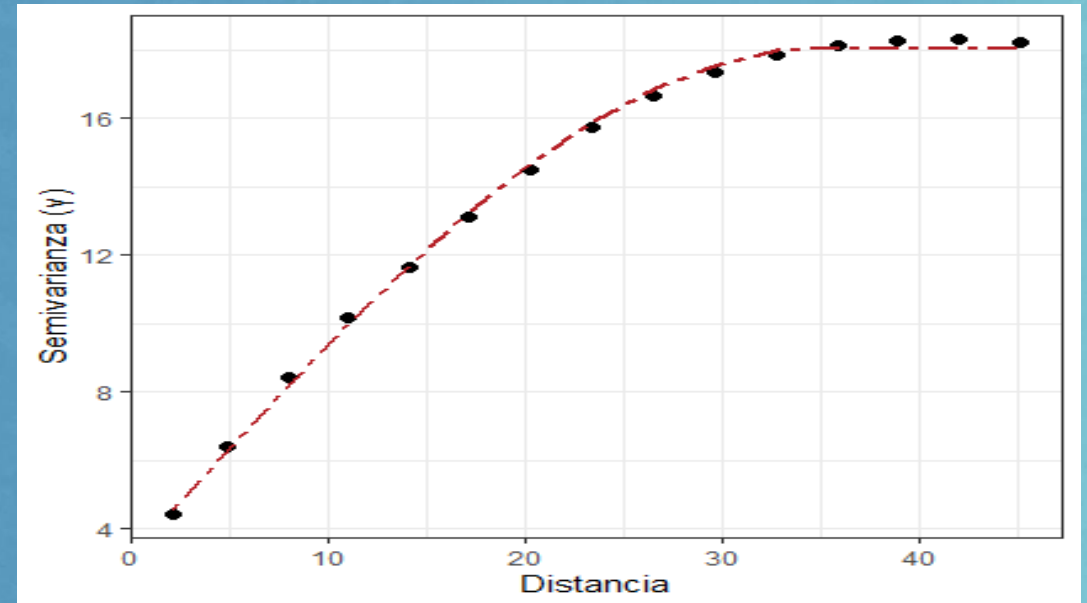
Con base en $\varepsilon(\mathbf{s})$ se realiza todo el ajuste de semivariogramas teóricos. Las opciones tenidas en cuenta para este paso son:

- Exponencial
- Esférico
- Gaussiano
- Mattern

En cada caso se evaluó el de mejor ajuste y este fue el que se conservó para pasos posteriores.

Este paso busca capturar la dependencia espacial, es decir, la correlación entre los valores de hectárea debido a su proximidad geográfica.

Una vez realizado el proceso de ajuste del componente espacial y de regresión se procede a recolectar las métricas definidas sobre la muestra de prueba.





¡Muchas Gracias!

obs_inmobiliario@igac.gov.co