



Universidad del  
**Rosario**

# Analisis Avanzado de Datos

## W6. Métodos de Suavización: Splines

FERNEY ALBERTO BELTRAN MOLINA

Escuela de Ingeniería, Ciencia y Tecnología

Matemáticas Aplicadas y Ciencias de la Computación

# Profesor

FERNEY ALBERTO BELTRAN MOLINA

[ferney.beltran@urosario.edu.co](mailto:ferney.beltran@urosario.edu.co)

Ingeniero Electrónico.

Magister en TIC

Candidato Doctor en TIC

Director del Centro de investigación e innovación CEINTECCI.

Miembro de la junta directiva Avanciencia

Procesamiento y análisis de datos basadas en IA.

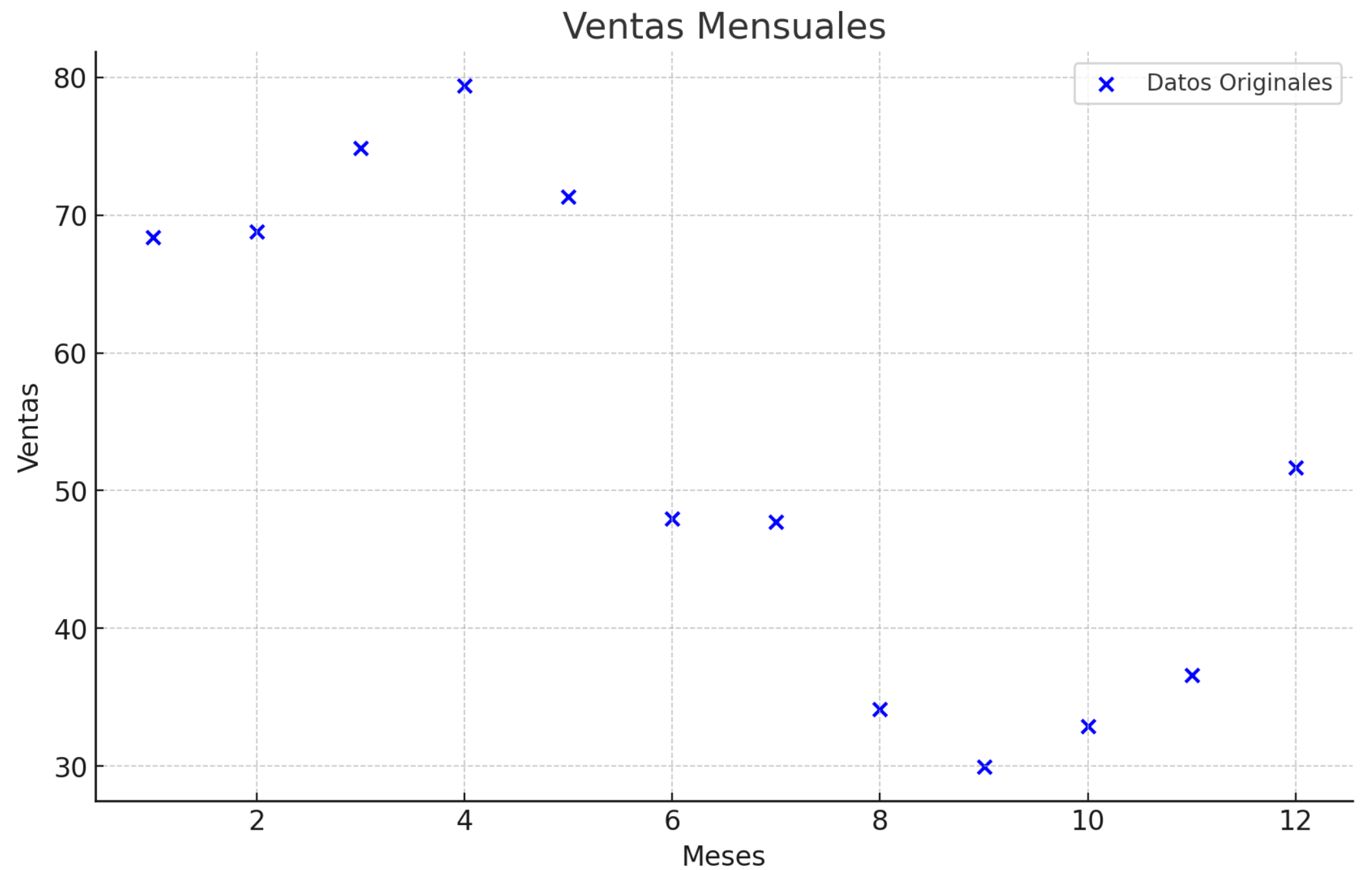
Simulación y modelado por computación,

Optimizan Sistemas de procesamiento en hardware y software

Diseño de sistemas electrónicos reconfigurables

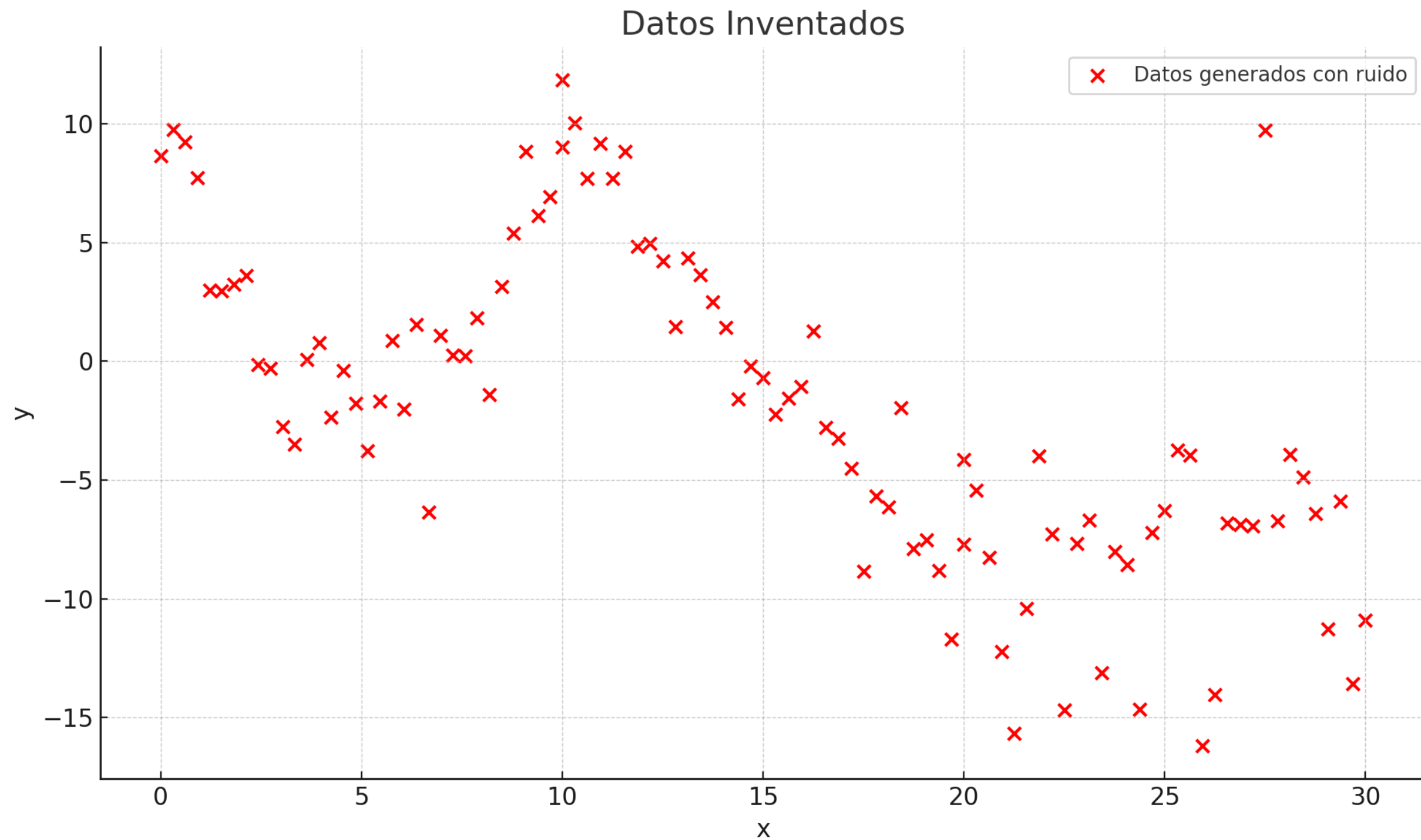
# Problema de regresión

- **Sobreajuste** (Overfitting)
- **Flexibilidad**

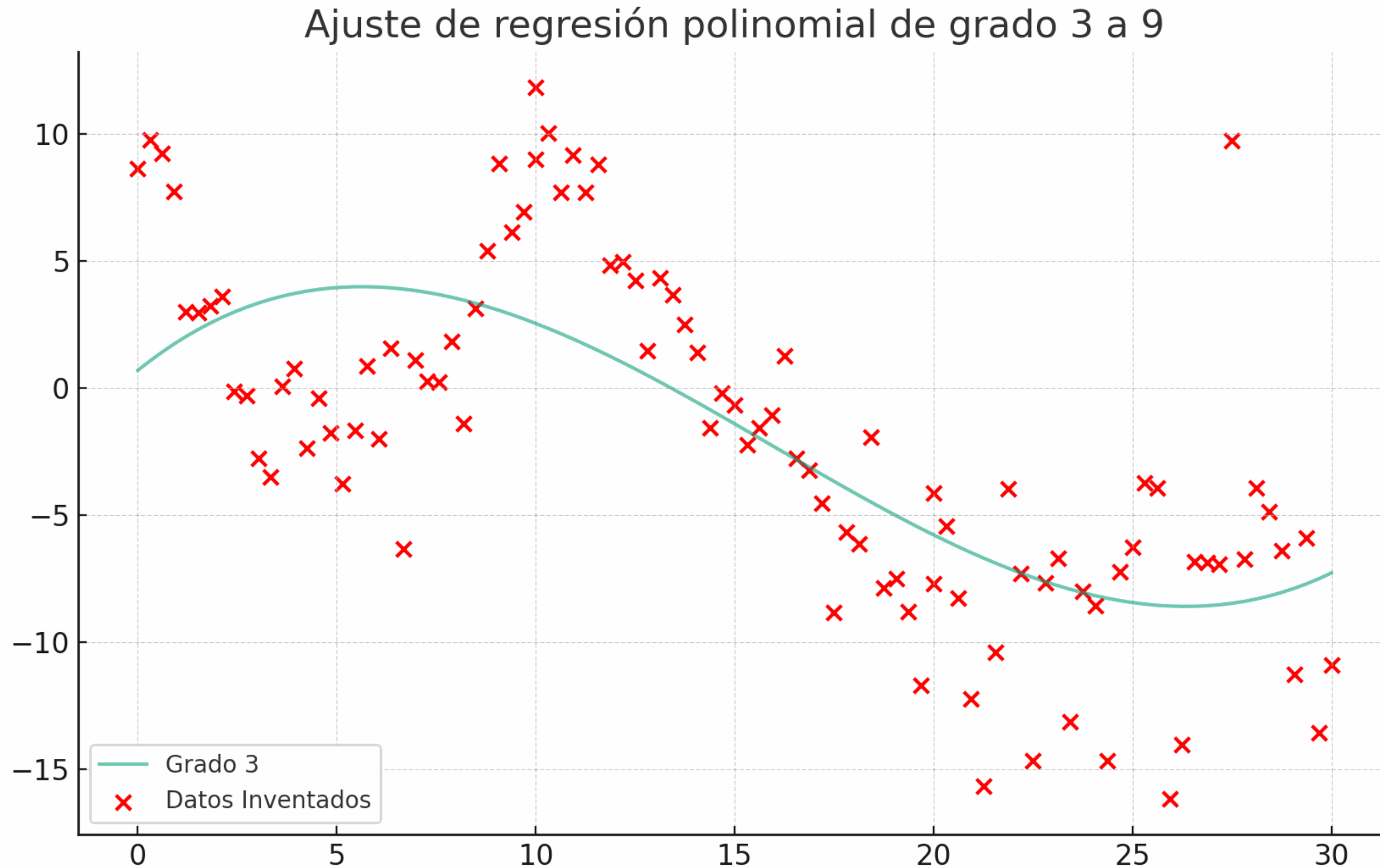


# Problema de regresión

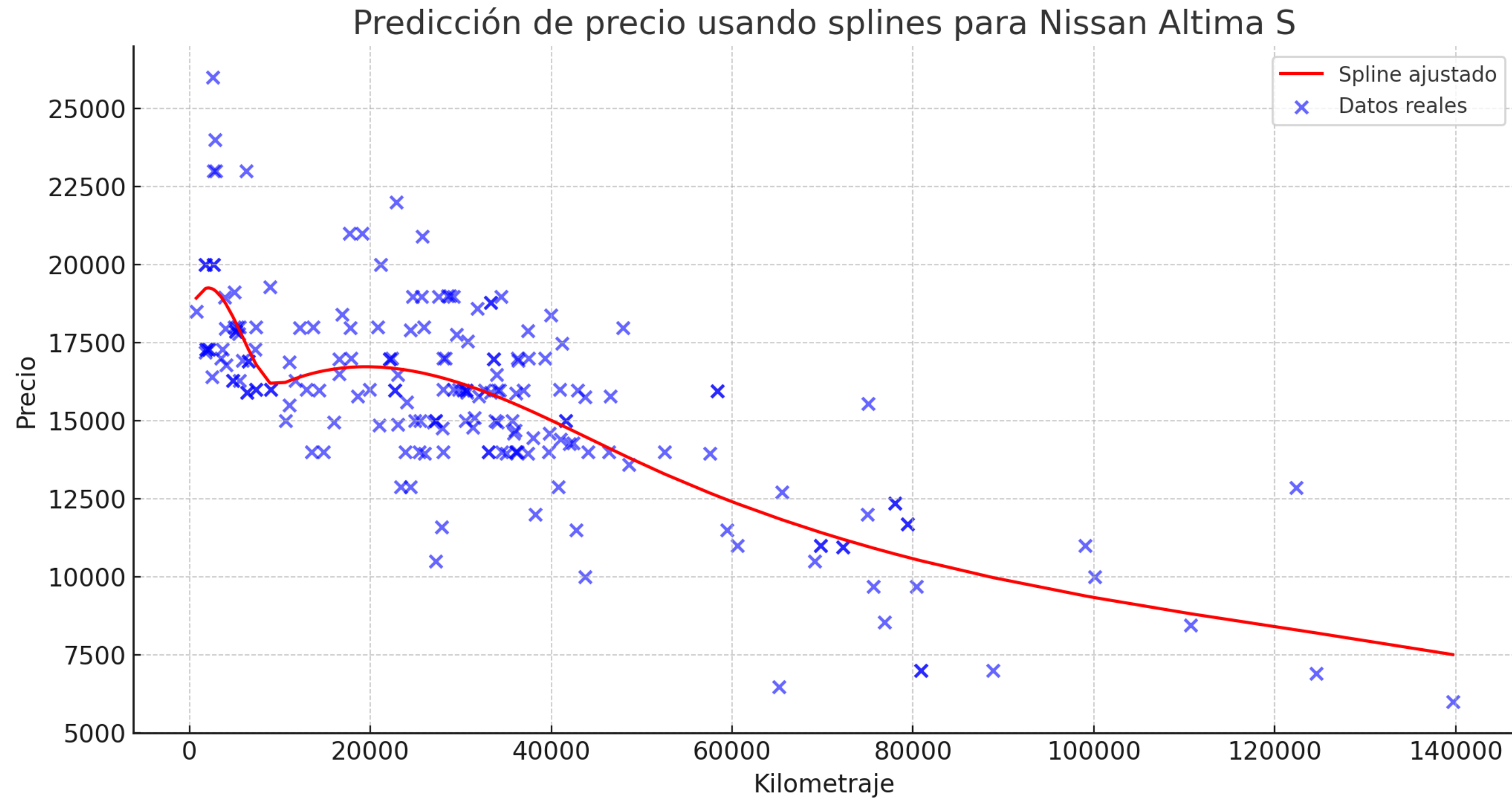
- **Sensibilidad a Variaciones Pequeñas**
- **Interpretabilidad**



# Cuál es el mejor polinomio

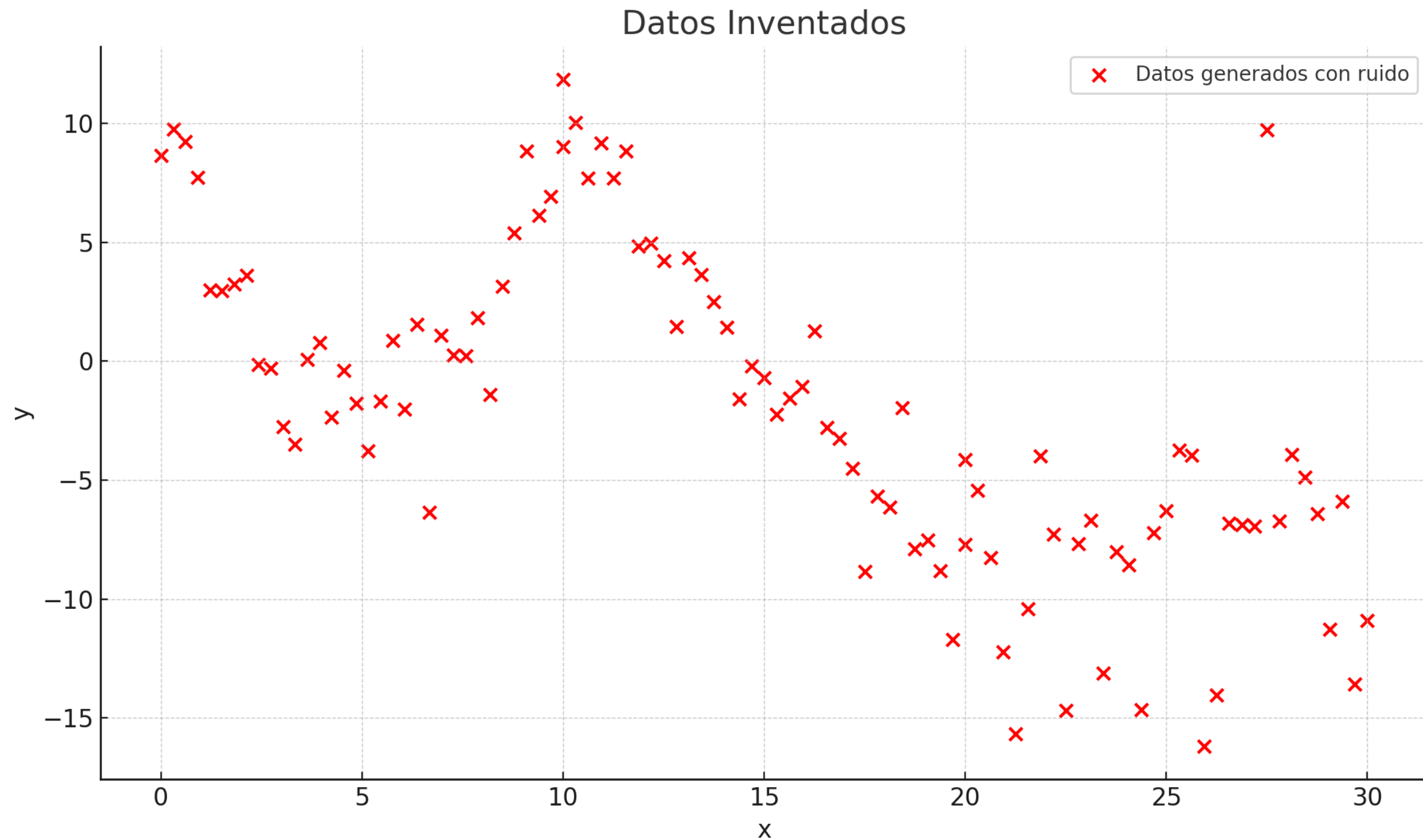


# Qué grado de polinomio es?

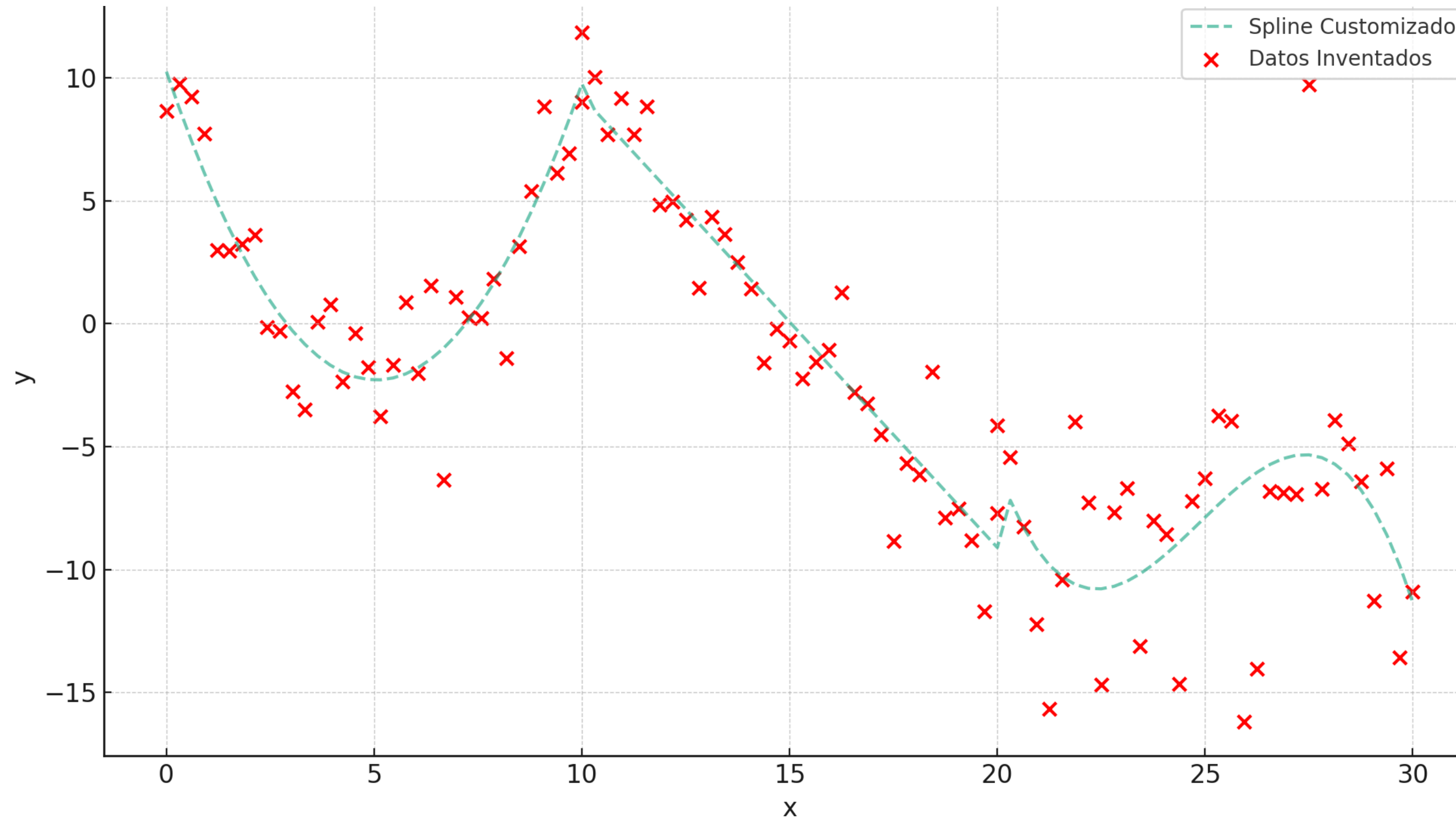


# Spline

objetivo dividir el rango del predictor  $X$  en  $K$  subintervalos. Para cada una de las nuevas regiones se ajusta una función polinómica, introduciendo una serie de restricciones que hacen que los extremos de cada función se aproximen a los de las funciones de las regiones colindantes.



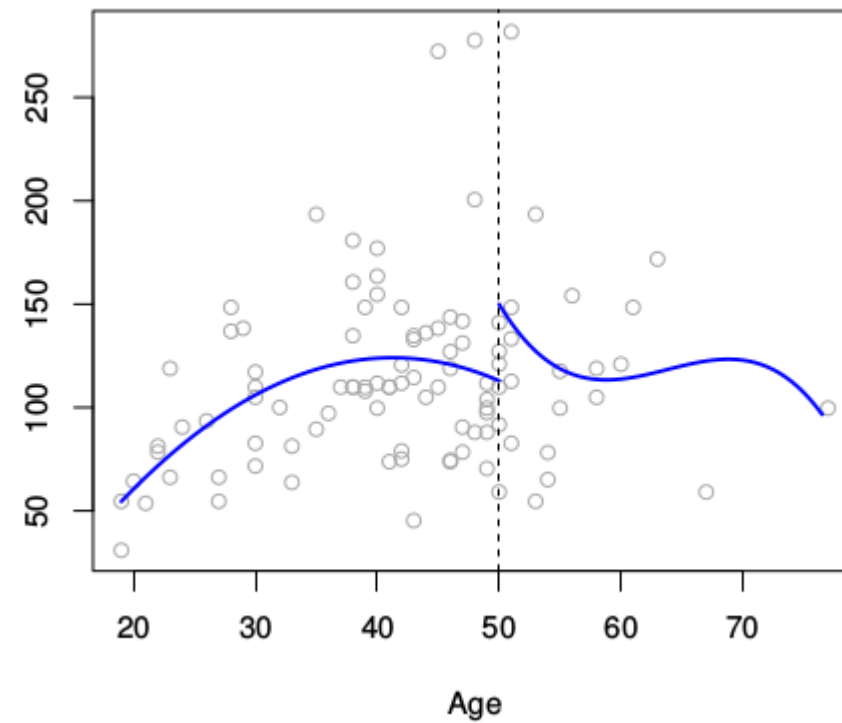
# Solución Que observamos



Discontinuidad



# Dividir

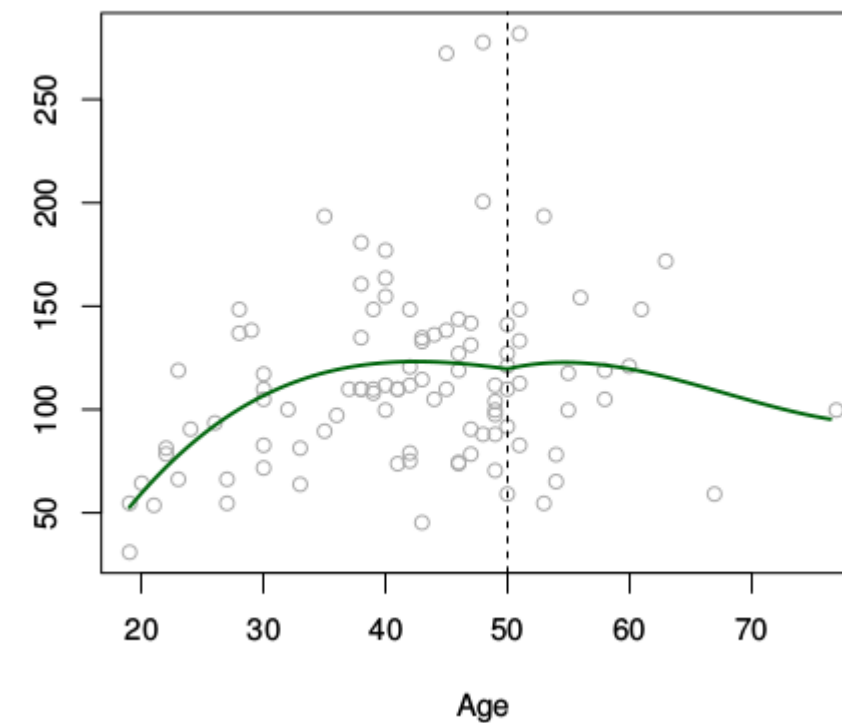


Se ajustan dos funciones polinómicas por mínimos cuadrados

$$y_i = \begin{cases} \beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2 + \beta_{31}x_i^3 & \text{si } x_i < c \\ \beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2 + \beta_{32}x_i^3 & \text{si } x_i \geq c \end{cases}$$

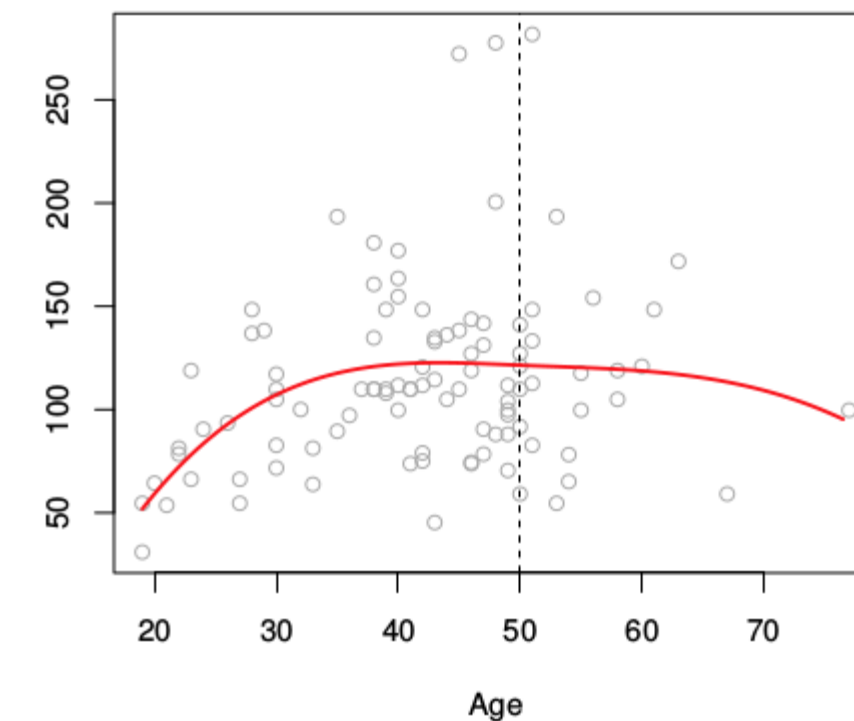
La flexibilidad de los modelos  
Se controla con los Knots

Mayor Knot mayor flexibilidad



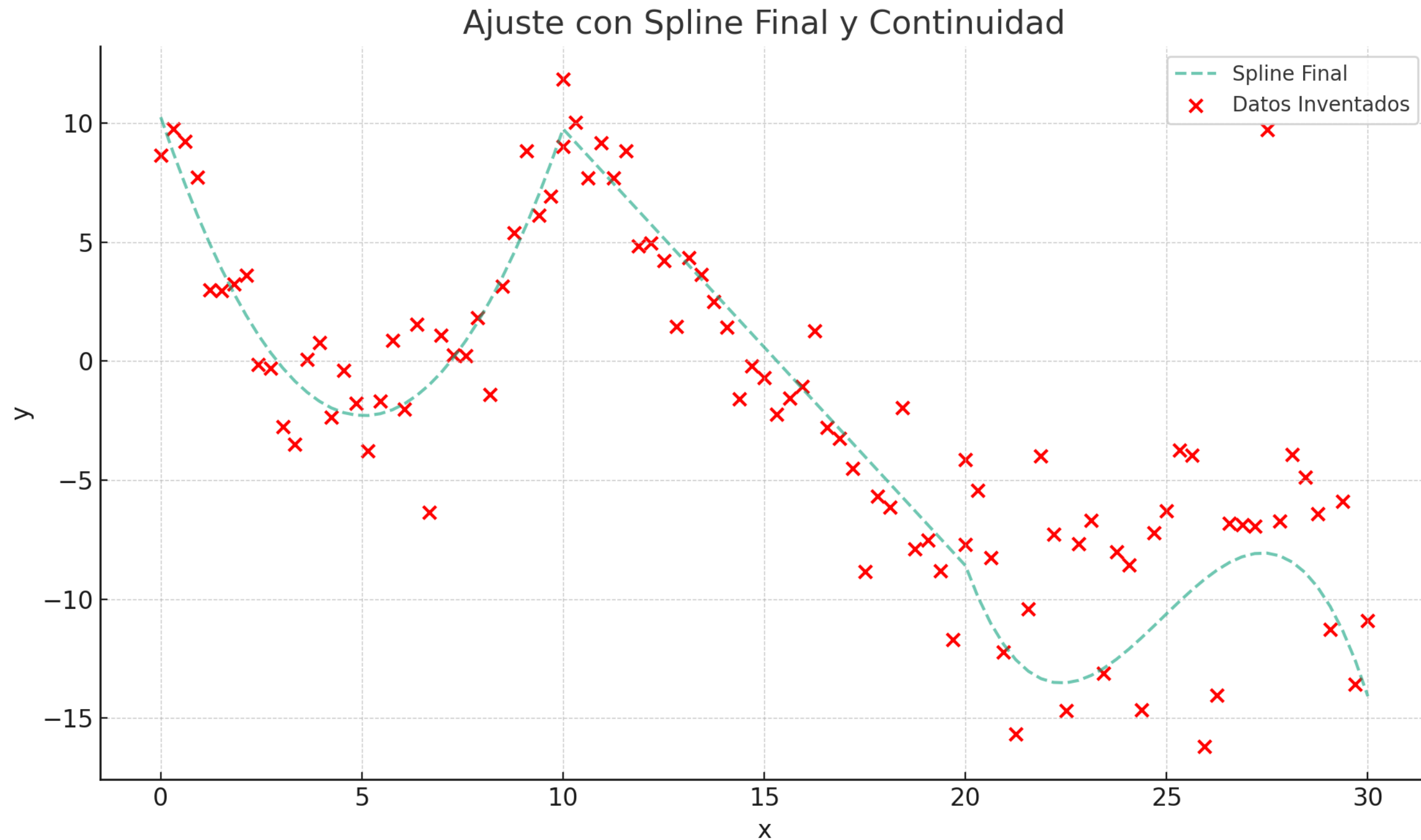
El polinomio de la región A termina en el  
mismo punto donde empieza el polinomio de  
la región B

Abrupto y poco natural



discontinuidad y exceso de flexibilidad

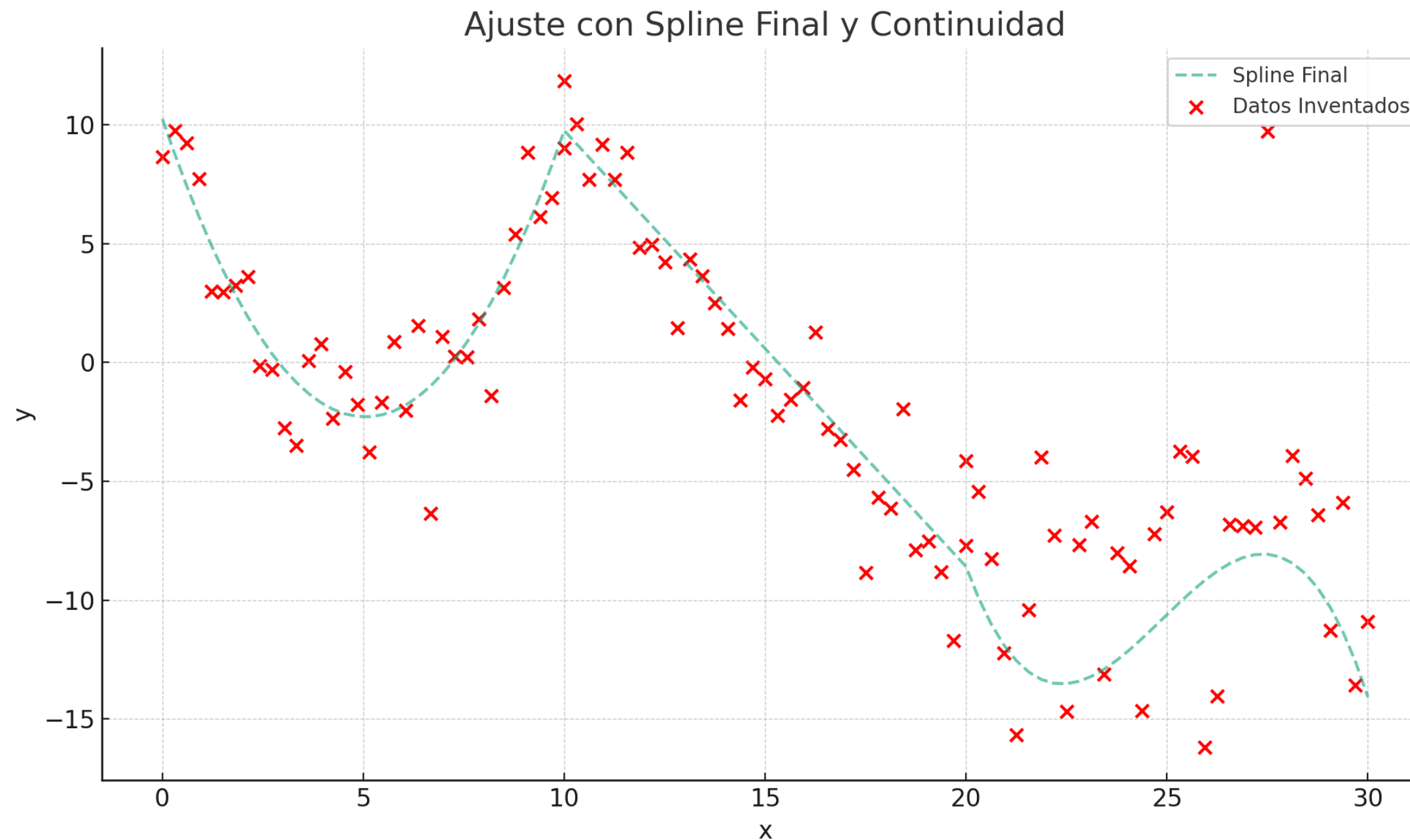
# Solución Que observamos



**Espaciado Uniforme**  
**Selección Manual**  
**Optimización**

son los puntos donde se conectan los diferentes  
segmentos polinómicos en un spline

# Solución Que observamos



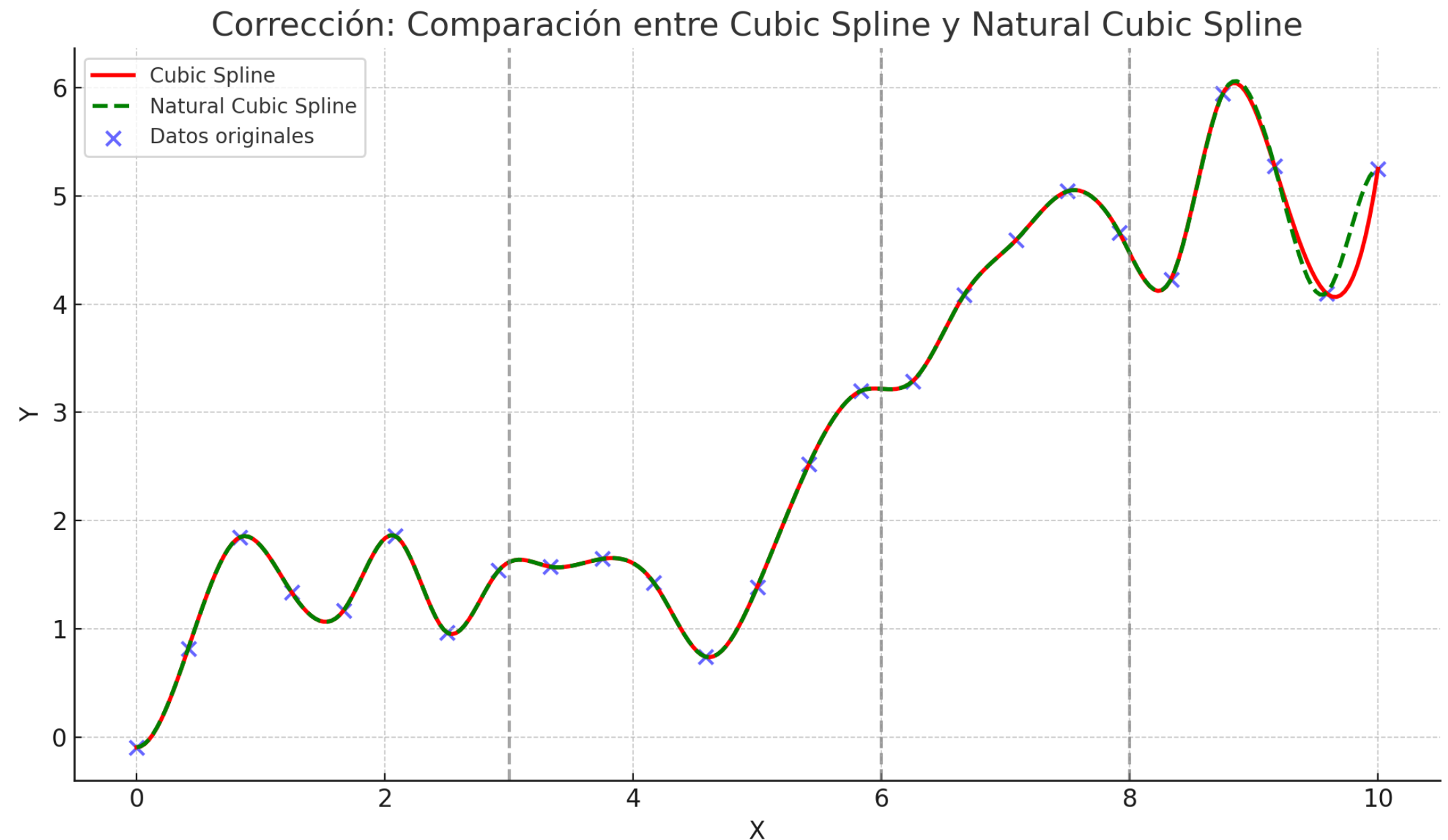
Las regresiones con splines pueden tener mucha varianza en los extremos superior e inferior del predictor, lo que genera intervalos de confianza muy amplios

Obligar al spline a ser lineal en los extremos, se evita que el modelo se ajuste demasiado a ruido o variaciones en esos extremos, lo que puede llevar a predicciones poco realistas fuera de los datos

# Cubic spline y natural spline

Cubic Spline: Ajusta polinomios cúbicos entre pares de datos, garantizando suavidad y continuidad.

Natural Cubic Spline: Es un cubic spline que se vuelve lineal en los extremos del dominio para reducir la variabilidad y el sobreajuste en esos puntos.

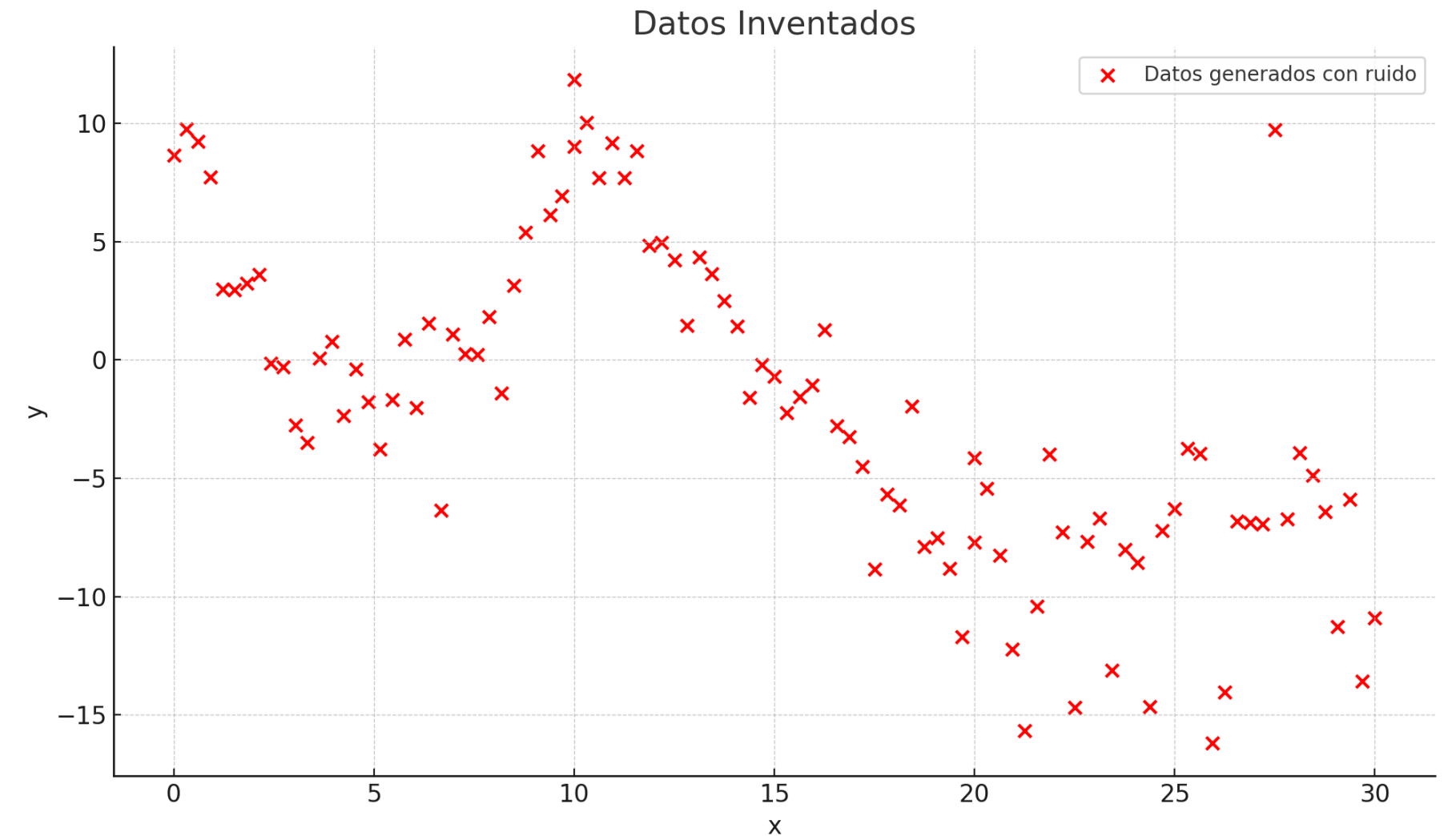


# Cubic spline

**Localidad:** A diferencia de un polinomio global que afecta a todos los datos, un spline afecta solo a un segmento local de los datos.

**Continuidad:** garantizamos que el spline sea continuo. las primeras y segundas derivadas sean continuas. Esto evita "codos" o discontinuidades en las uniones, conocidas como "nodos" o "knots".

**Flexibilidad:** Dado que ajustamos diferentes polinomios a diferentes segmentos, los splines pueden adaptarse a funciones más complejas que un solo polinomio global.

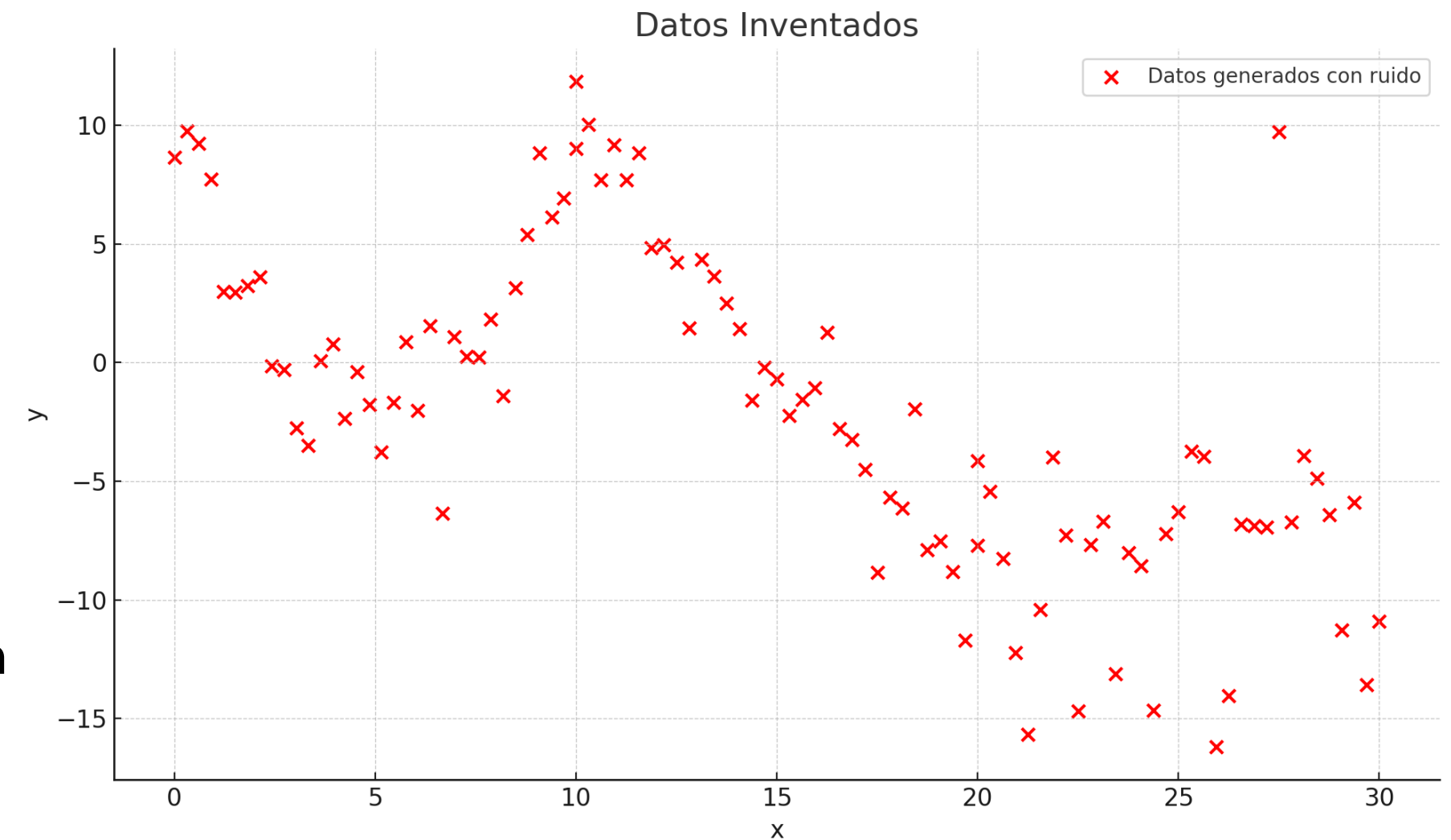


# Natural spline

**Linealidad en los extremos:** La función resultante del spline es lineal antes del primer nodo y después del último nodo. Esto significa que fuera de estos puntos, el spline se asemeja a una línea recta.

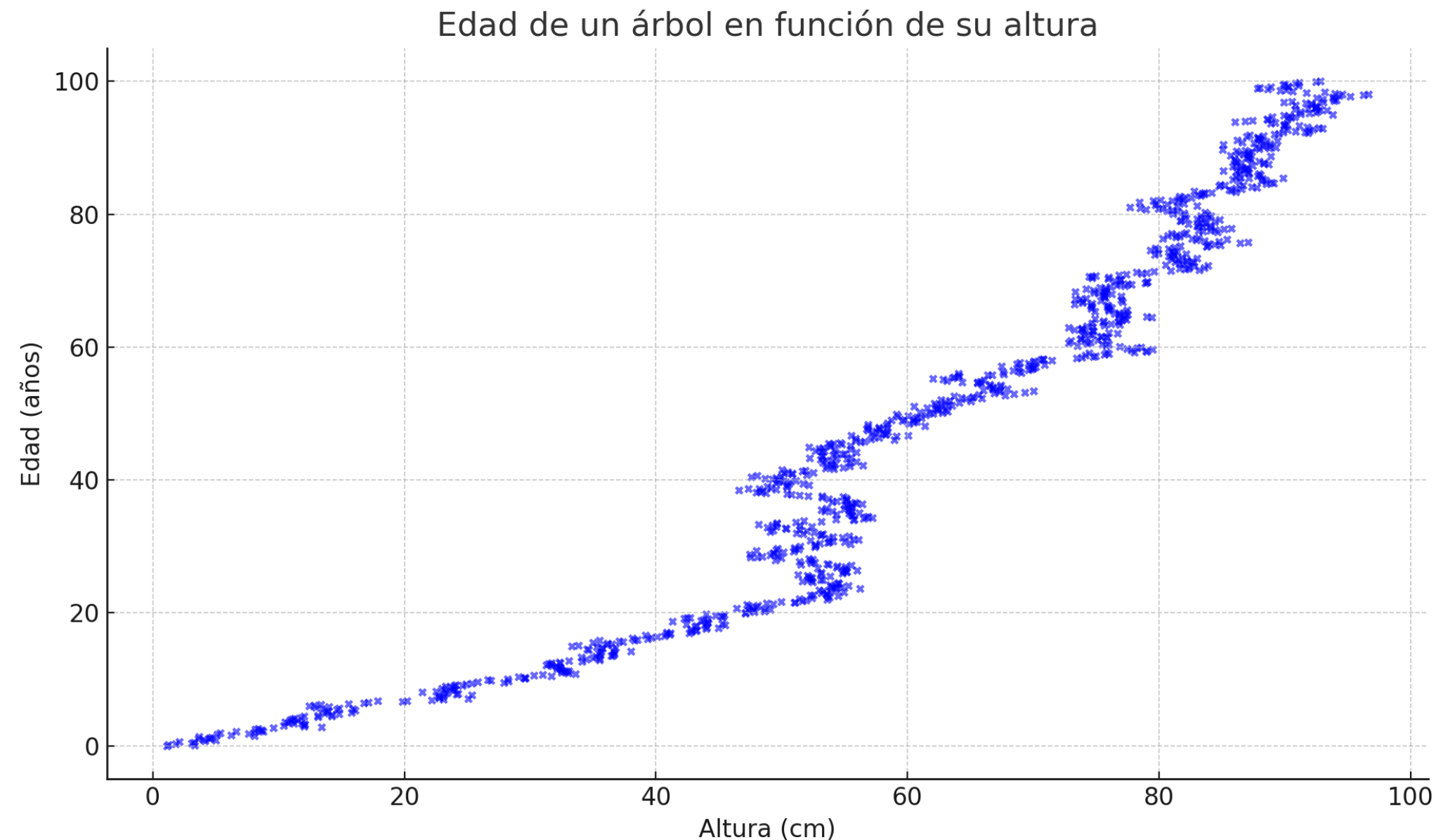
**Reducción de la varianza en los extremos:** Los splines cúbicos regulares pueden tener una alta varianza en los extremos. Al forzar al spline a ser lineal en los extremos, reducimos esta varianza.

**Menos parámetros** Debido a las restricciones adicionales, un natural spline generalmente tiene menos parámetros para estimar que un spline cúbico regular..



# Ejercicio 1

Imagina que tienes datos sobre el crecimiento de un árbol y quieres modelar su la edad función de su altura. Si simplemente ajustas un polinomio, podría tener problemas en las edades extremas (muy jóvenes o muy viejas). Un spline cúbico permitiría un ajuste más flexible en diferentes segmentos de edad. Un natural spline garantizaría que este ajuste no sea demasiado errático en las edades extremas, proporcionando una transición suave hacia una línea recta.



## Ejercicio 2

Toma uno de los tres ejercicios de la semana de evaluación y mejorar la calidad del modelo usando spline

Puedes usar cualquier librería  
scipy o statsmodels patsy

Siente libre de explorar

<https://patsy.readthedocs.io/en/latest/spline-regression.html>

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.UnivariateSpline.html>

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.LSQUnivariateSpline.html>



# Qué grado de polinomio es?

