UTADEO
UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO

Curso Métodos y Modelos

Profesora:
Karen Ballesteros-González PhD.

Modelos Basados en Datos

Modelos Estadísticos

i. Regresión lineal y no lineal (Ej. predicción de tendencias)
ii. Análisis de series temporales (Ej. modelos para pronóstico de demanda)

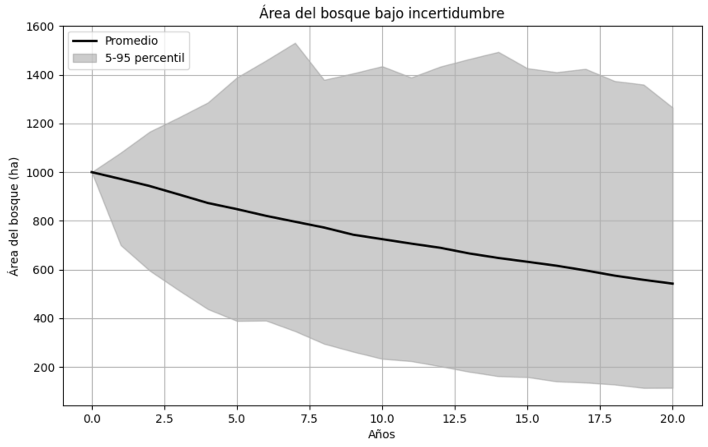
Objetivo:
Introducir y aplicar modelos estadísticos y de aprendizaje automático para análisis y predicción de datos, incluyendo regression y series temporales.

Ejercicio: Modelo Estocástico de Reforestación y Captura de Carbono

Este ejercicio simula el crecimiento del área de un bosque reforestado durante 20 años, considerando la posibilidad de eventos aleatorios que afectan dicho crecimiento.

Descripción del modelo
Cada año, el área del bosque puede crecer o reducirse dependiendo de un evento aleatorio:

- **Incendio** (probabilidad = 0.2): reduce el área en un 30%.
- **Plaga** (probabilidad = 0.1): reduce el área en un 15%.
- **Protección** (probabilidad = 0.3): aumenta la tasa de crecimiento a 8%.
- **Sin evento** (probabilidad = 0.4): el crecimiento se mantiene en 5%.

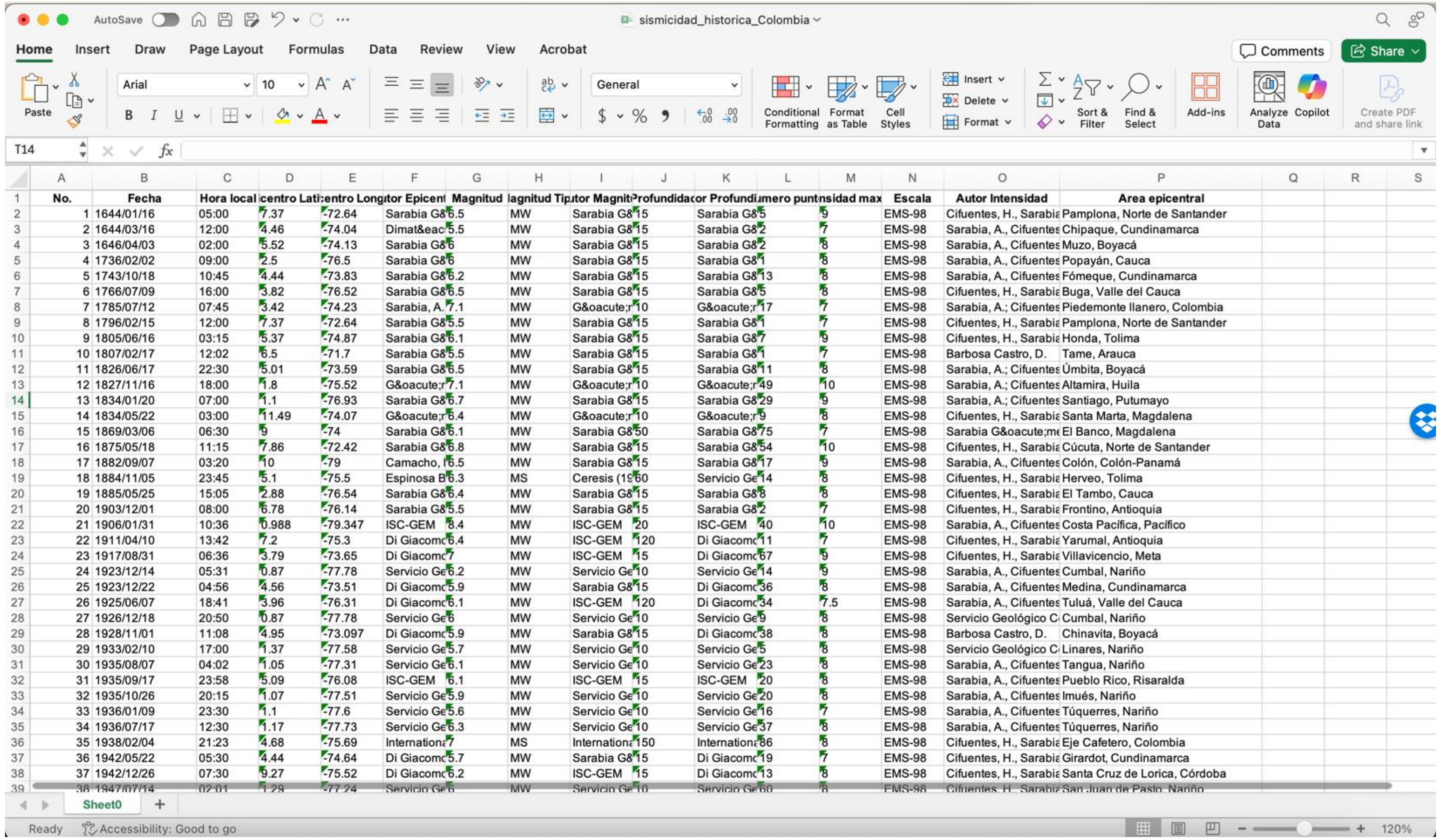


La ecuación básica de crecimiento de bosque es:

$$A_t = A_{t-1} \times (1 + r_t)$$

Donde:
A_t: área del bosque en el año (t)
r_t: tasa de crecimiento o decrecimiento, según el evento ocurrido en el año (t)

Ejercicio: Cadenas de Markov y Probabilidad de Sismos



No.	Fecha	Hora local	centro Lati	centro Long	tor Epicent	Magnitud	lagnitud Tip	tor Magnit	Profundidac	or Profundi	umero punt	nsidad max	Escala	Autor Intensidad	Area epicentral
1	1644/01/16	05:00	7.37	-72.64	Sarabia G&	6.5	MW	Sarabia G&	15	Sarabia G&	5	9	EMS-98	Cifuentes, H., Sarabia	Pamplona, Norte de Santander
2	1644/03/16	12:00	4.46	-74.04	Dimat&eac	5.5	MW	Sarabia G&	15	Sarabia G&	2	7	EMS-98	Sarabia, A., Cifuentes	Chipaque, Cundinamarca
3	1646/04/03	02:00	5.52	-74.13	Sarabia G&	6	MW	Sarabia G&	15	Sarabia G&	2	8	EMS-98	Sarabia, A., Cifuentes	Muzo, Boyacá
4	1736/02/02	09:00	2.5	-76.5	Sarabia G&	6	MW	Sarabia G&	15	Sarabia G&	1	8	EMS-98	Sarabia, A., Cifuentes	Popayán, Cauca
5	1743/10/18	10:45	4.44	-73.83	Sarabia G&	6.2	MW	Sarabia G&	15	Sarabia G&	13	8	EMS-98	Sarabia, A., Cifuentes	Fómeque, Cundinamarca
6	1766/07/09	16:00	3.82	-76.52	Sarabia G&	6.5	MW	Sarabia G&	15	Sarabia G&	5	8	EMS-98	Cifuentes, H., Sarabia	Buga, Valle del Cauca
7	1785/07/12	07:45	3.42	-74.23	Sarabia, A.	7.1	MW	Gór	10	Gór	17	7	EMS-98	Sarabia, A.; Cifuentes	Piedemonte llanero, Colombia
8	1796/02/15	12:00	7.37	-72.64	Sarabia G&	5.5	MW	Sarabia G&	15	Sarabia G&	1	7	EMS-98	Cifuentes, H., Sarabia	Pamplona, Norte de Santander
9	1805/06/16	03:15	5.37	-74.87	Sarabia G&	6.1	MW	Sarabia G&	15	Sarabia G&	7	9	EMS-98	Cifuentes, H., Sarabia	Honda, Tolima
10	1807/02/17	12:02	6.5	-71.7	Sarabia G&	5.5	MW	Sarabia G&	15	Sarabia G&	1	7	EMS-98	Barbosa Castro, D.	Tame, Arauca
11	1826/06/17	22:30	5.01	-73.59	Sarabia G&	6.5	MW	Sarabia G&	15	Sarabia G&	11	8	EMS-98	Sarabia, A.; Cifuentes	Úmbita, Boyacá
12	1827/11/16	18:00	1.8	-75.52	Gór	7.1	MW	Gór	10	Gór	49	10	EMS-98	Sarabia, A.; Cifuentes	Altamira, Huila
13	1834/01/20	07:00	1.1	-76.93	Sarabia G&	6.7	MW	Sarabia G&	15	Sarabia G&	29	9	EMS-98	Sarabia, A.; Cifuentes	Santiago, Putumayo
14	1834/05/22	03:00	11.49	-74.07	Gór	6.4	MW	Gór	10	Gór	9	8	EMS-98	Cifuentes, H., Sarabia	Santa Marta, Magdalena
15	1869/03/06	06:30	9	-74	Sarabia G&	6.1	MW	Sarabia G&	50	Sarabia G&	75	7	EMS-98	Sarabia Góme	El Banco, Magdalena
16	1875/05/18	11:15	7.86	-72.42	Sarabia G&	6.8	MW	Sarabia G&	15	Sarabia G&	54	10	EMS-98	Cifuentes, H., Sarabia	Cúcuta, Norte de Santander
17	1882/09/07	03:20	10	-79	Camacho, I	6.5	MW	Sarabia G&	15	Sarabia G&	17	9	EMS-98	Sarabia, A., Cifuentes	Colón, Colón-Panamá
18	1884/11/05	23:45	5.1	-75.5	Espinosa B	6.3	MS	Ceresis (19	60	Servicio Ge	14	8	EMS-98	Cifuentes, H., Sarabia	Herveo, Tolima
19	1885/05/25	15:05	2.88	-76.54	Sarabia G&	6.4	MW	Sarabia G&	15	Sarabia G&	8	8	EMS-98	Cifuentes, H., Sarabia	El Tambo, Cauca
20	1903/12/01	08:00	6.78	-76.14	Sarabia G&	5.5	MW	Sarabia G&	15	Sarabia G&	2	7	EMS-98	Cifuentes, H., Sarabia	Frontino, Antioquia
21	1906/01/31	10:36	0.988	-79.347	ISC-GEM	8.4	MW	ISC-GEM	20	ISC-GEM	40	10	EMS-98	Sarabia, A., Cifuentes	Costa Pacífica, Pacífico
22	1911/04/10	13:42	7.2	-75.3	Di Giacomo	6.4	MW	ISC-GEM	120	Di Giacomo	11	7	EMS-98	Cifuentes, H., Sarabia	Yarumal, Antioquia
23	1917/08/31	06:36	3.79	-73.65	Di Giacomo	7	MW	ISC-GEM	15	Di Giacomo	67	9	EMS-98	Cifuentes, H., Sarabia	Villavicencio, Meta
24	1923/12/14	05:31	0.87	-77.78	Servicio Ge	6.2	MW	Servicio Ge	10	Servicio Ge	14	9	EMS-98	Sarabia, A., Cifuentes	Cumbal, Nariño
25	1923/12/22	04:56	4.56	-73.51	Di Giacomo	5.9	MW	Sarabia G&	15	Di Giacomo	36	8	EMS-98	Sarabia, A., Cifuentes	Medina, Cundinamarca
26	1925/06/07	18:41	3.96	-76.31	Di Giacomo	6.1	MW	ISC-GEM	120	Di Giacomo	34	7.5	EMS-98	Sarabia, A., Cifuentes	Tuluá, Valle del Cauca
27	1926/12/18	20:50	0.87	-77.78	Servicio Ge	6	MW	Servicio Ge	10	Servicio Ge	9	8	EMS-98	Servicio Geológico C	Cumbal, Nariño
28	1928/11/01	11:08	4.95	-73.097	Di Giacomo	5.9	MW	Sarabia G&	15	Di Giacomo	38	8	EMS-98	Barbosa Castro, D.	Chinavita, Boyacá
29	1933/02/10	17:00	1.37	-77.58	Servicio Ge	5.7	MW	Servicio Ge	10	Servicio Ge	5	8	EMS-98	Servicio Geológico C	Linares, Nariño
30	1935/08/07	04:02	1.05	-77.31	Servicio Ge	6.1	MW	Servicio Ge	10	Servicio Ge	23	8	EMS-98	Sarabia, A., Cifuentes	Tangua, Nariño
31	1935/09/17	23:58	5.09	-76.08	ISC-GEM	6.1	MW	ISC-GEM	15	ISC-GEM	20	8	EMS-98	Sarabia, A., Cifuentes	Pueblo Rico, Risaralda
32	1935/10/26	20:15	1.07	-77.51	Servicio Ge	5.9	MW	Servicio Ge	10	Servicio Ge	20	8	EMS-98	Sarabia, A., Cifuentes	Imués, Nariño
33	1936/01/09	23:30	1.1	-77.6	Servicio Ge	5.6	MW	Servicio Ge	10	Servicio Ge	16	7	EMS-98	Sarabia, A., Cifuentes	Túquerres, Nariño
34	1936/07/17	12:30	1.17	-77.73	Servicio Ge	6.3	MW	Servicio Ge	10	Servicio Ge	37	8	EMS-98	Sarabia, A., Cifuentes	Túquerres, Nariño
35	1938/02/04	21:23	4.68	-75.69	Internationa	7	MS	Internationa	150	Internationa	86	8	EMS-98	Cifuentes, H., Sarabia	Eje Cafetero, Colombia
36	1942/05/22	05:30	4.44	-74.64	Di Giacomo	5.7	MW	Sarabia G&	15	Di Giacomo	19	7	EMS-98	Cifuentes, H., Sarabia	Girardot, Cundinamarca
37	1942/12/26	07:30	9.27	-75.52	Di Giacomo	6.2	MW	ISC-GEM	15	Di Giacomo	13	8	EMS-98	Cifuentes, H., Sarabia	Santa Cruz de Lorica, Córdoba
38	1947/07/14	02:01	1.29	-77.24	Servicio Ge	6	MW	Servicio Ge	10	Servicio Ge	60	8	EMS-98	Cifuentes, H., Sarabia	San Juan de Pasto, Nariño

https://sish.sgc.gov.co/visor/sesionServlet?metodo=irAEpicentrosTodos&idDepartamento=&idMunicipio=&cuadranteXMin=&cuadranteXMax=&cuadranteYMin=&cuadranteYMax=

Modelos Estadísticos: Regresión lineal y no lineal

a). Regresión lineal y no lineal

Breve introducción teórica:

- Modelo de regresión lineal simple y múltiple.
- Extensión a modelos no lineales: polinómica, exponencial.

Aplicación práctica:

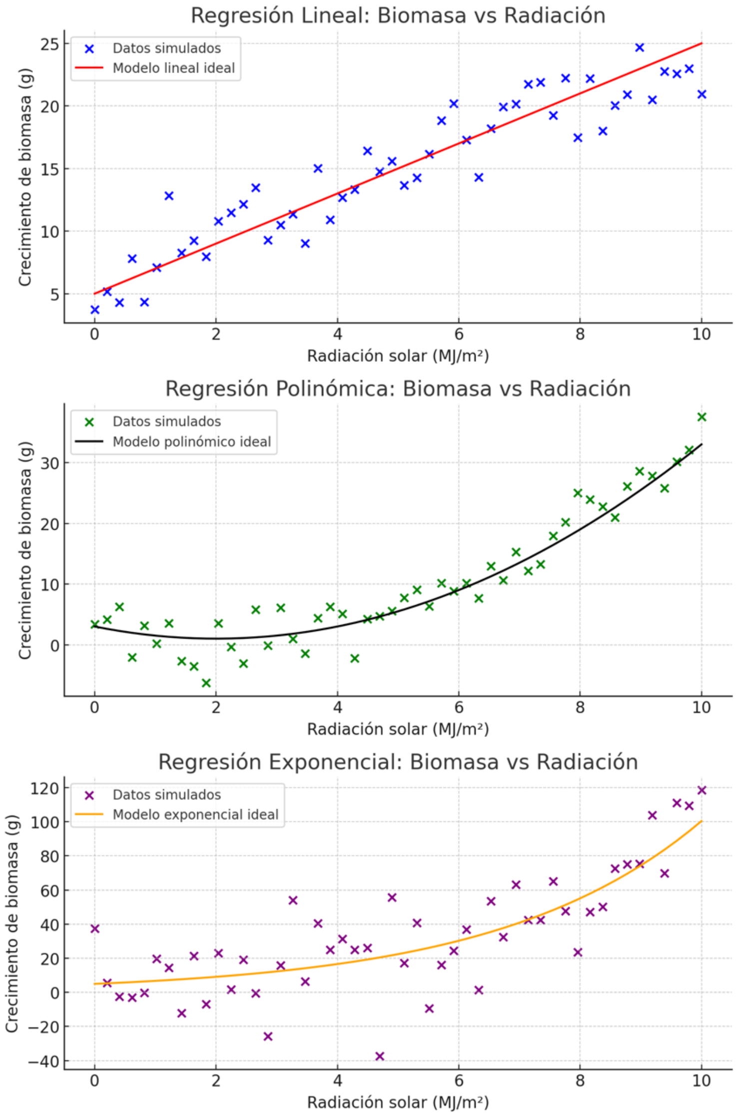
- Predecir tendencia de temperatura/precipitación o calidad del aire con datos históricos.

Modelos Estadísticos: Regresión lineal y no lineal

¿Qué es la regresión?

La regresión es una técnica estadística que permite modelar y analizar la relación entre una variable dependiente (*respuesta*) y una o más variables independientes (*predictoras*). Su objetivo principal es predecir valores de la variable dependiente basándose en el comportamiento de las independientes.

- Regresión Lineal Simple
- Regresión Lineal Multiple
- Regresión No Lineal
    - R. Polinómica
    - R. Exponencial



Modelos Estadísticos: Regresión lineal y no lineal

Regresión Lineal Simple

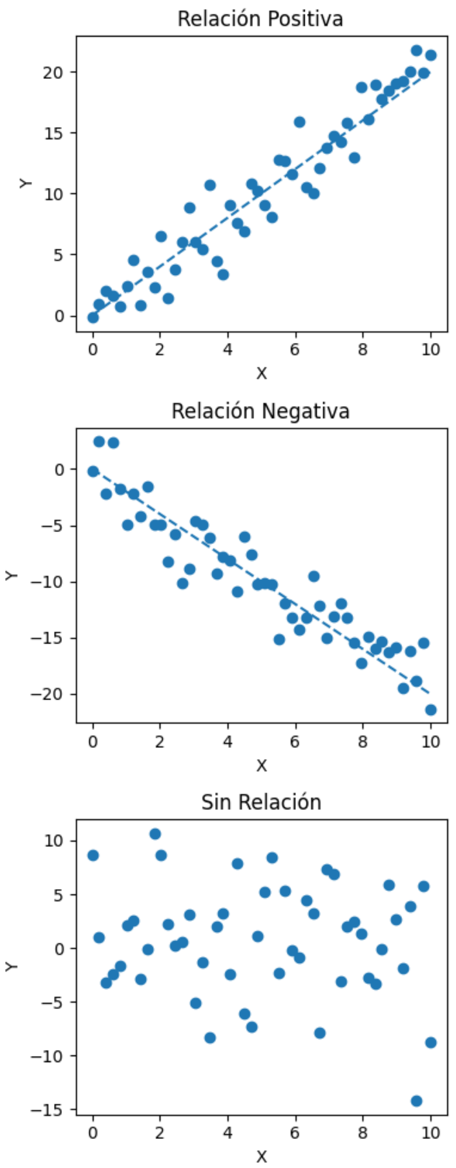
Es un modelo matemático que describe una relación lineal entre dos variables:

Modelo:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y: variable dependiente
x: variable independiente
β_0: intercepto
β_1: pendiente (efecto del cambio de x sobre y)
ε: término de error aleatorio

Se utiliza cuando se busca predecir una variable en función de una sola variable explicativa.



Modelos Estadísticos: Regresión lineal y no lineal

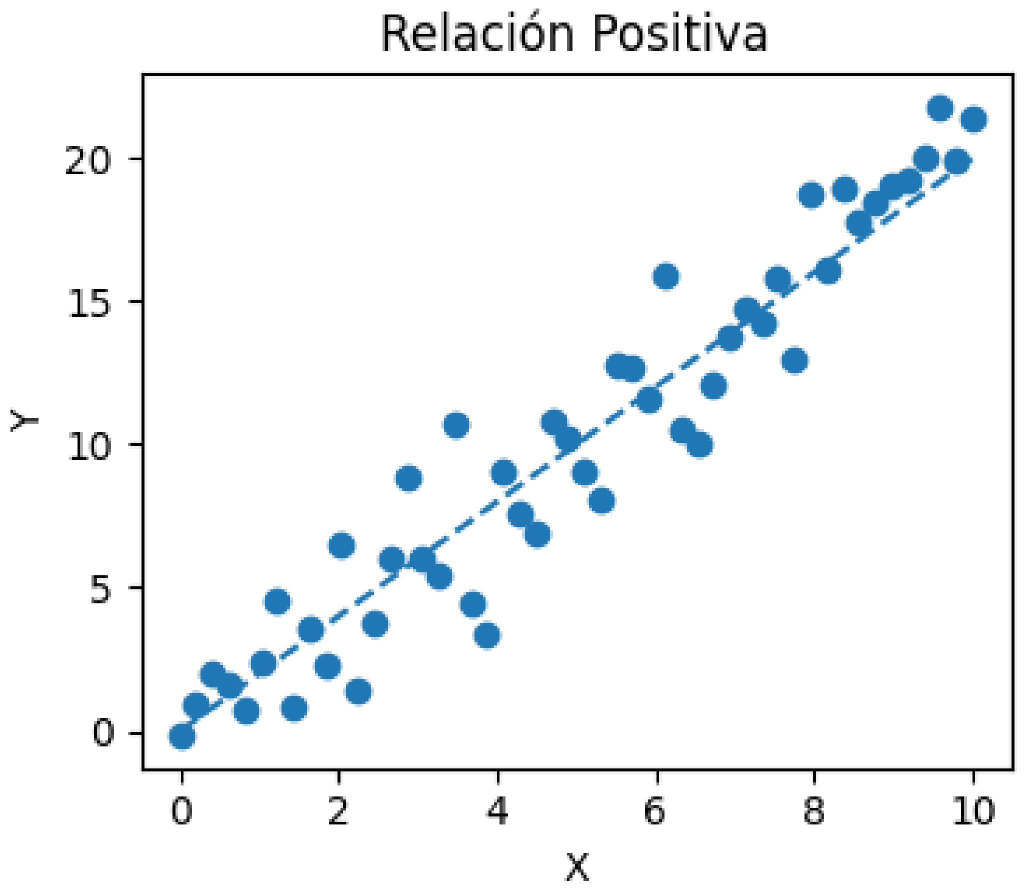
Intercepto β_0 : Valor de y cuando $x = 0$.

$$\beta_0 = \bar{y} + \beta_1 \bar{x}$$

Pendiente β_1: Tasa de cambio de y por cada unidad de x.

$$\beta_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Error ε: Diferencia de cada punto a la línea del modelo. Línea que minimiza el error.



Modelos Estadísticos: Regresión lineal y no lineal

Predicción de caudal de un río en función del clima:

Se desea predecir el caudal diario (Q) de un río con base en variables meteorológicas. Se dispone de registros históricos de:

$$Q = 0.5 * P + 0.3 * T + 0.2 * R + \varepsilon$$

Q: Caudal simulado (en m^3/s).
P: Precipitación diaria (en mm).
T: Temperatura media diaria (en °C).
R: Radiación solar diaria (en MJ/m^2).
$\varepsilon = np.random.normal(0, 5, size = n)$: Término de error aleatorio con distribución normal de media 0 y desviación estándar 5, que introduce variabilidad para simular datos más realistas.

Interpretación de los coeficientes

- **0.5**: Por cada milímetro adicional de precipitación, se espera un aumento de 0.5 m^3/s en el caudal, manteniendo constantes las demás variables.
- **0.3**: Cada grado Celsius adicional en la temperatura media diaria se asocia con un incremento de 0.3 m^3/s en el caudal, bajo las mismas condiciones.
- **0.2**: Un aumento de 1 MJ/m^2 en la radiación solar diaria se relaciona con un incremento de 0.2 m^3/s en el caudal, considerando constantes las otras variables.

*Nota: **La formula de caudal, es una expresión utilizada para simular datos de caudal (Q) en función de variables climáticas: precipitación (P), temperatura (T) y radiación solar (R). Esta fórmula no proviene de un estudio específico, sino que se emplea para generar datos sintéticos con el propósito de practicar modelos de regresión lineal múltiple en un contexto ambiental.*

Modelos Estadísticos: Regresión lineal y no lineal

Regresión Lineal Multiple

Modelo:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

Permite predecir una variable en función de varias variables explicativas $x_1, x_2, \ldots, x_p$. Es útil cuando los fenómenos son complejos y dependen de múltiples factores.

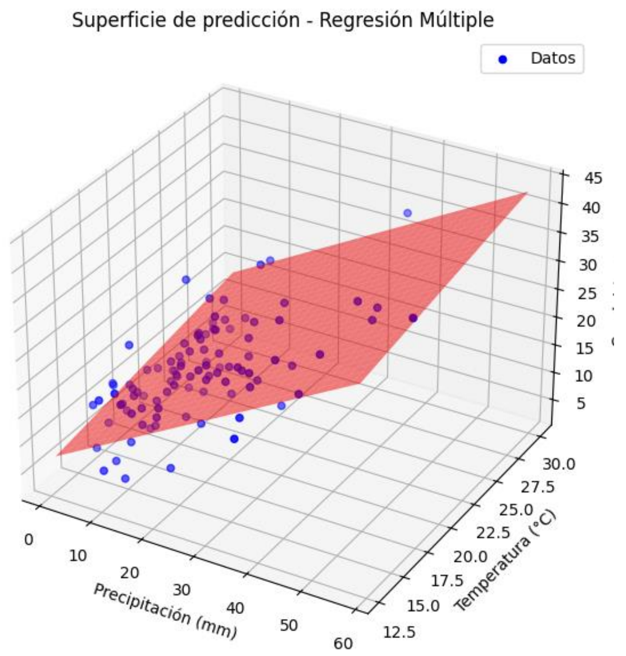
- Usamos **más de una variable independiente**.
- Ideal para sistemas complejos que dependen de múltiples factores.

Ejemplo:

$$Caudal = \beta 0 + \beta 1 \cdot Precipitación + \beta 2 \cdot Temperatura + \beta 3 \cdot Radiación + \varepsilon$$

Donde:

- $\boldsymbol{y}$**:** vector columna con los valores observados de la variable dependiente
- $\boldsymbol{X}$**:** matriz de diseño con las variables explicativas (incluye columna de unos para el intercepto)
- $\boldsymbol{\beta}$: vector de coeficientes $[\beta_0, \beta_1, \ldots, \beta_p]$
- **ε:** vector de errores (diferencia entre observado y predicho)



Extensiones: Regresión No Lineal

Es un modelo en el que la relación entre la variable dependiente y y las variables independientes x **no puede representarse como una combinación lineal simple** de los parámetros.

- En una regresión **lineal**, la relación es una recta.
- En una regresión **no lineal**, puede ser una curva:

a) Regresión Polinómica:

- Extiende la regresión lineal simple añadiendo potencias del predictor.
- Modelo con términos de orden superior.
- Permite capturar curvaturas en los datos.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon$$

Modelos Estadísticos: Regresión lineal y no lineal

a) Regresión Polinómica:

- Extiende la regresión lineal simple **añadiendo potencias del predictor.**
- Modelo con términos de orden superior.
- Permite capturar curvaturas en los datos.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon$$

Ventajas

- Captura relaciones más complejas que la recta.
- Se ajusta con herramientas de regresión lineal.
- Fácil de visualizar e interpretar hasta grado 2 o 3.

Desventajas:

- Usar grados muy altos puede llevar a **sobreajuste** (la curva sigue demasiado los datos, incluyendo el ruido).
- La interpretación de los coeficientes se vuelve menos intuitiva.

Ejemplo: Si tienes precipitación x y caudal y, puedes ajustar:

Regresión lineal:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Regresión polinómica de grado 2:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

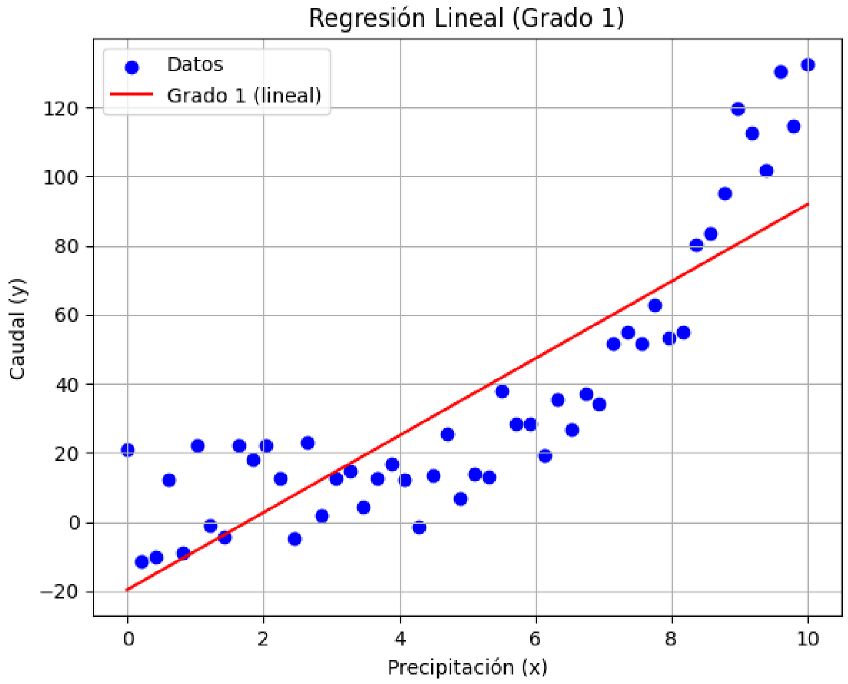
Grado 3:

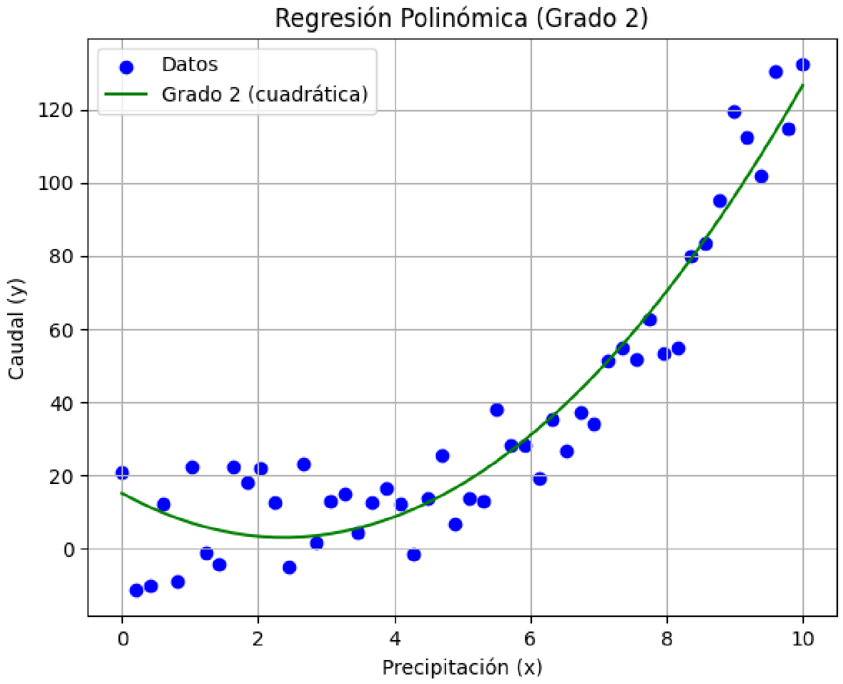
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

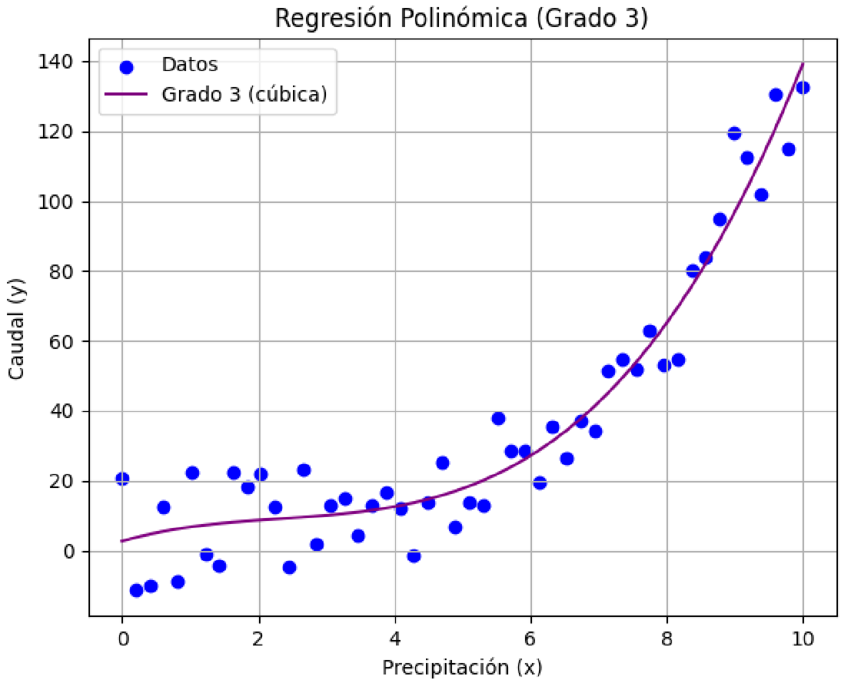
Mientras más alto el grado, más flexible la curva, pero también puede sobreajustar - **overfitting**.

Modelos Estadísticos: Regresión lineal y no lineal

Regresión No Lineal - Regresión Polinómica







Regresión Lineal (Grado 1)

- La recta trata de ajustar los datos con una única pendiente.
- No capta la curvatura de la relación.

Regresión Polinómica (Grado 2)

- La curva cuadrática (una parábola) comienza a ajustarse mejor a la forma de los datos.
- Ideal cuando la relación es simétrica y tiene un único mínimo o máximo.

Regresión Polinómica (Grado 3)

- Modelo más flexible, capaz de adaptarse a cambios de curvatura.
- Puede captar mejor relaciones más complejas, pero también corre más riesgo de sobreajuste.

Modelos Estadísticos: Regresión lineal y no lineal

b) Regresión Exponencial:

Cuando el comportamiento entre variables sigue una forma **no lineal** y es muy útil cuando los datos siguen un patrón de **crecimiento o decaimiento acelerado**. :

Ejemplo exponencial:

$$y = \beta_0 e^{\beta_1 x}$$

Este modelo no se puede ajustar directamente con regresión lineal, pero podemos **linealizarlo** tomando logaritmos:

$$\log(y) = \log(\beta_0) + \beta_1 x$$

$y \ = \ \log(y)$
$\beta_0 = \log(\beta_0)$
$\beta_1 = \beta_1$

Modelos Estadísticos: Regresión lineal y no lineal

Paso 5: Regresión exponencial

Pregunta: ¿El caudal responde de forma exponencial a la precipitación?

```python
import numpy as np
from sklearn.linear_model import LinearRegression

# Supongamos que tienes tus datos:
x = df[['Precipitacion_mm']]
y = df['Caudal_m3s']

# Transformar y aplicando logaritmo natural
y_log = np.log(np.where(y > 0, y, 0.1))  # para evitar log(0)

# Ajustar modelo lineal sobre log(y)
modelo_exp = LinearRegression().fit(x, y_log)

# Coeficientes:
print("log(a):", modelo_exp.intercept_)
print("b:", modelo_exp.coef_[0])
```

```
log(a): 2.2564735985253934
b: 0.030212994632717347
```

Contexto del modelo

Se trata de un modelo de regresión lineal que busca explicar la relación entre la precipitación y el caudal de un río. Pero en lugar de predecir directamente el caudal (y), se modela el logaritmo natural del caudal (log(y)), lo cual implica que estamos asumiendo una relación exponencial entre ambas variables.

Regresión: ¿Cómo evaluamos si un modelo es bueno?

Métricas de Error en Modelos de Regresión

Se utilizan para evaluar el desempeño de un modelo de regresión.

- MAE (Error Absoluto Medio)
- MSE (Error Cuadrático Medio)
- RMSE (Raíz del Error Cuadrático Medio)
- R^2 (Coeficiente de Determinación)

Modelos de Regresión: Métricas de Error en Modelos de Regresión

MAE – Error Absoluto Medio

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

- Promedio del error absoluto.
- Fácil de interpretar.
- No penaliza errores grandes.

RMSE – Raíz del Error Cuadrático Medio

$$RMSE = \sqrt{MSE}$$

- MSE Similar al MSE, pero con las mismas unidades que y.
- Útil cuando se quiere interpretar el error en las mismas magnitudes del fenómeno.

MSE – Error Cuadrático Medio

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Promedio del cuadrado del error.
- Penaliza errores grandes más que el MAE.
- Muy usado en algoritmos de entrenamiento.

R^2 – Coeficiente de Determinación

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

- Mide la proporción de la varianza explicada por el modelo.
- Valores cercanos a 1 indican buen ajuste.
- Puede ser negativo si el modelo es peor que usar el promedio.

Modelos de Regresión: Métricas de Error en Modelos de Regresión

Métrica	**Fórmula**	**Qué mide**	**Interpretación**
R^2	$1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$	Varianza explicada por el modelo	$1 =$ perfecto, $0 =$ nulo
MSE	$\frac{1}{n}\sum(y_i - \hat{y}_i)^2$	Error cuadrático medio	Penaliza errores grandes
RMSE	$\sqrt{MSE}$	Raíz del MSE, en unidades reales	Fácil de interpretar
MAE	$\frac{1}{n}\sum\lvert y_i - \hat{y}_i\rvert$	Error absoluto medio	Robusto ante valores atípicos

Table 1: Métricas comunes para evaluar modelos de regresión

Modelos de Regresión: Ejericicio

Modelos Estadísticos: Análisis de series temporales

b). Análisis de series temporales

Teoría:

- Tendencia, estacionalidad, ruido.

Aplicación práctica:

- Predecir demanda de energía, lluvia mensual, $PM_{2.5}$.

Modelos Estadísticos: Análisis de series temporales

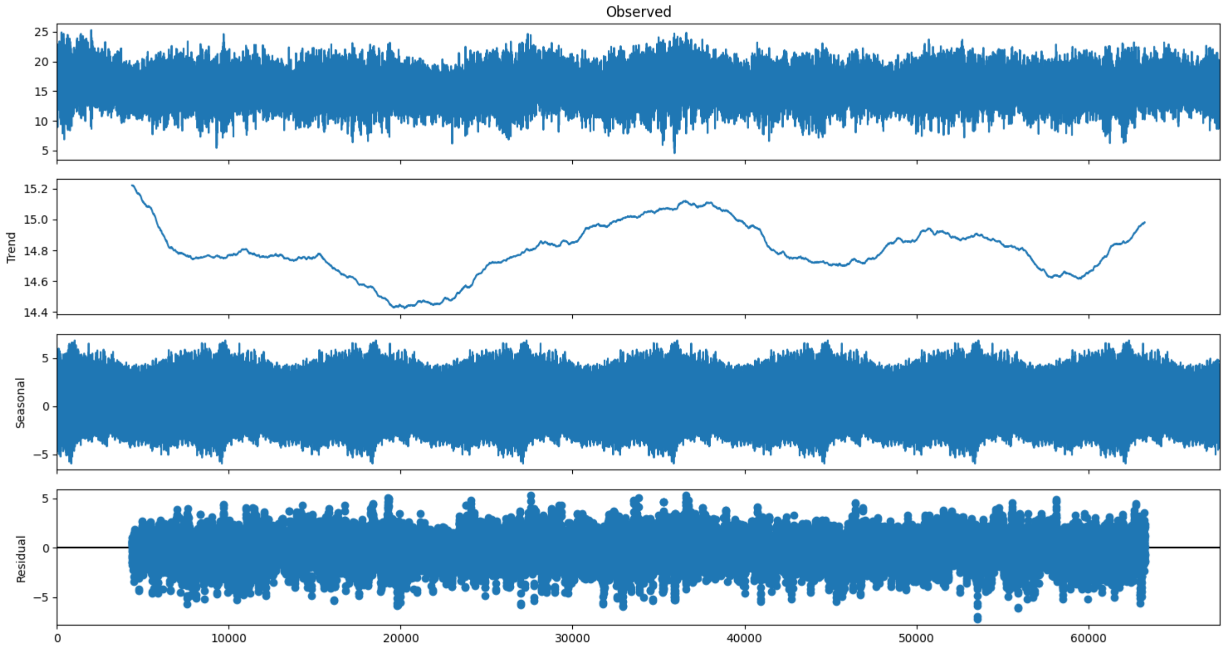
¿Qué es una serie temporal?

Es una secuencia de datos recolectados a lo largo del tiempo, con una frecuencia constante (horaria, diaria, mensual, etc.).

Componentes clave:

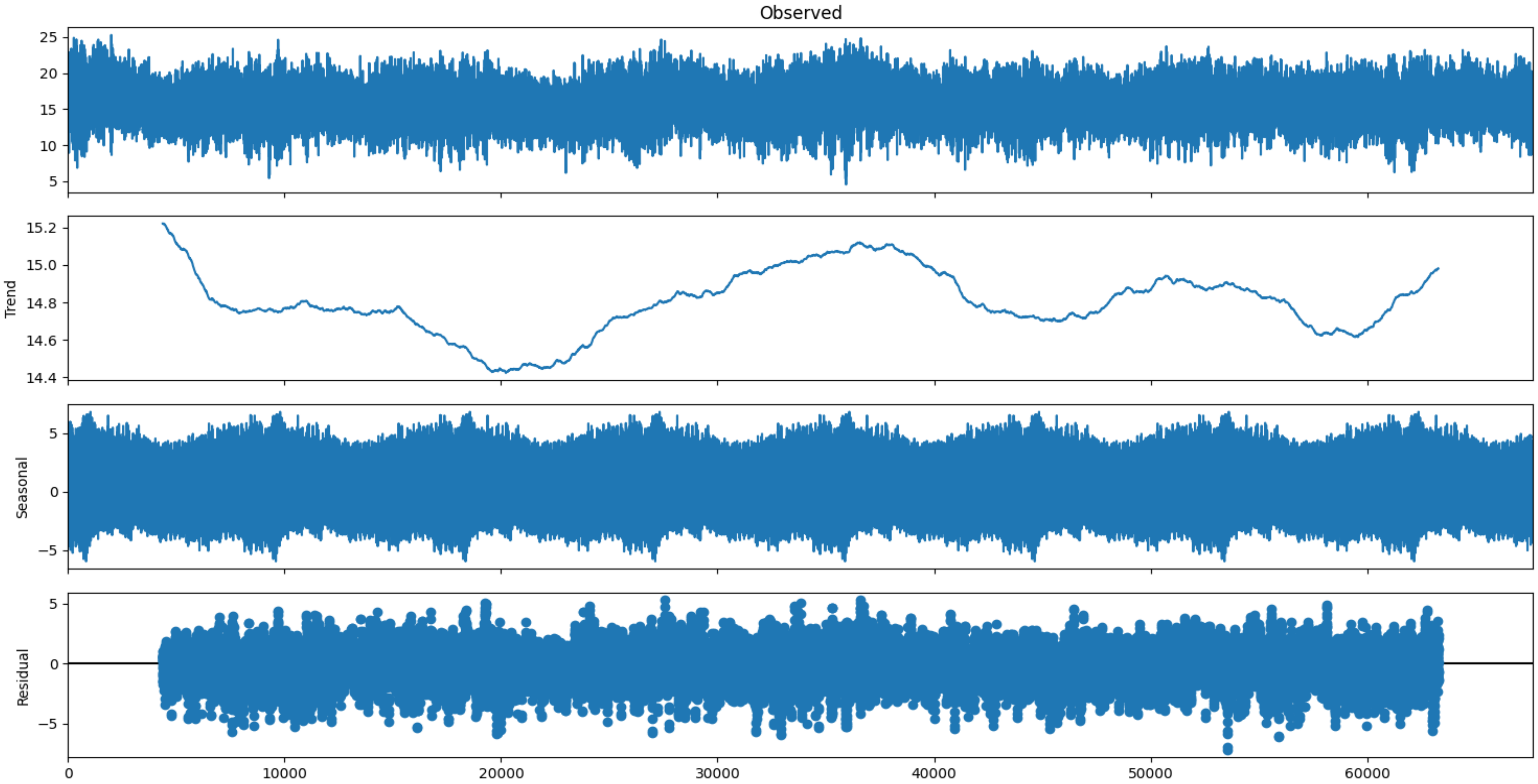
- **Tendencia (T):** cambio a largo plazo en el nivel de la serie.
- **Estacionalidad (S):** patrones que se repiten en intervalos fijos de tiempo (por ejemplo, estacionalidad mensual o diaria).
- **Ruido (R):** componente aleatorio no explicable por los anteriores.

```
seasonal_decompose()
```



Modelos Estadísticos: Análisis de series temporales

```
seasonal_decompose(promedio_hora.values, model='additive', period=8760)
```



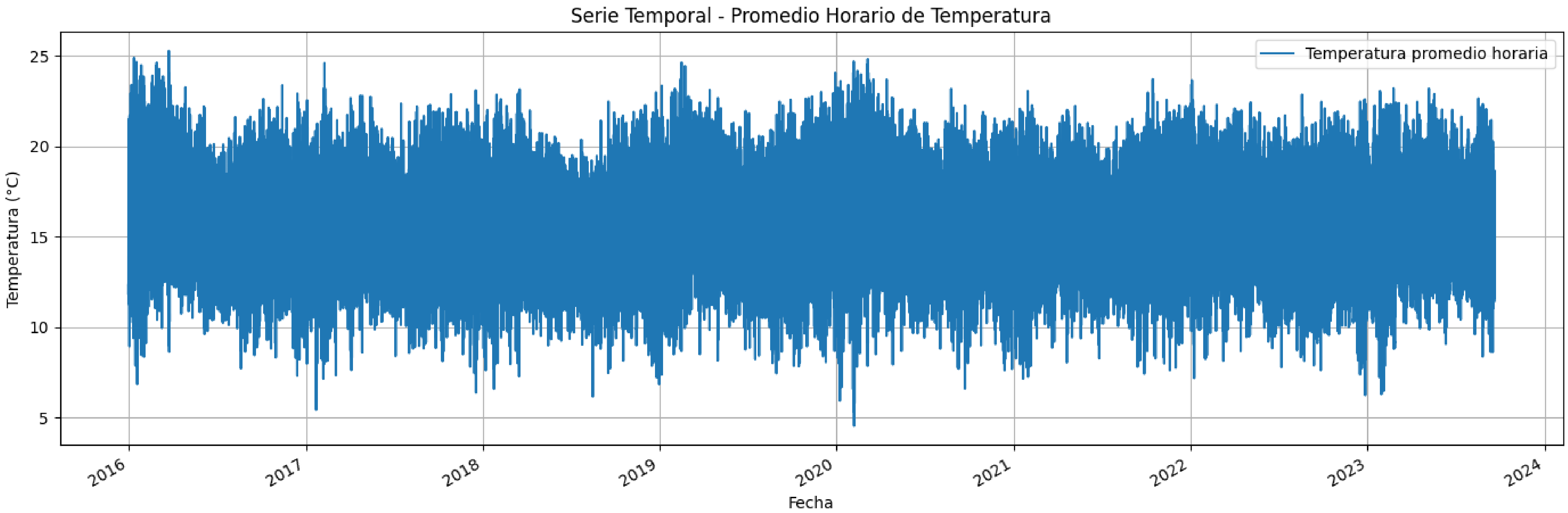
Modelos Estadísticos: Análisis de series temporales

Datos Observados – Serie de Tiempo

Modelo Aditivo de Serie Temporal

$$Yt = Tt + St + Rt$$

Yt = valor observado en el tiempo t
Tt = componente de **tendencia**
St = componente de **estacionalidad**
Rt = componente de **ruido o residuo**



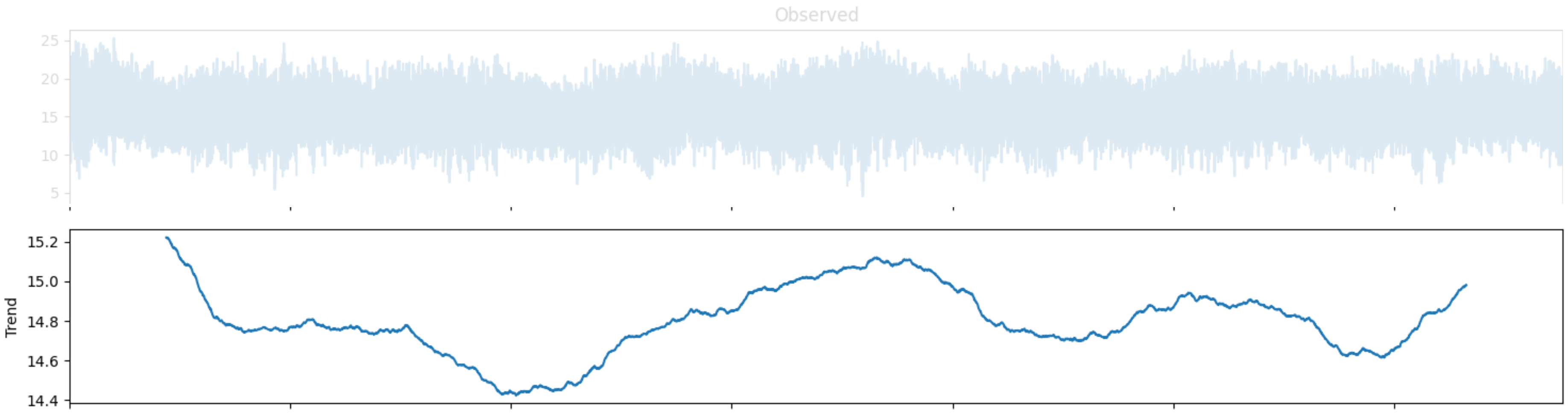
Modelos Estadísticos: Análisis de series temporales

Tendencia (T)

Es el **comportamiento general a largo plazo**. Puede ser creciente, decreciente o constante. Refleja cambios estructurales duraderos en el tiempo.

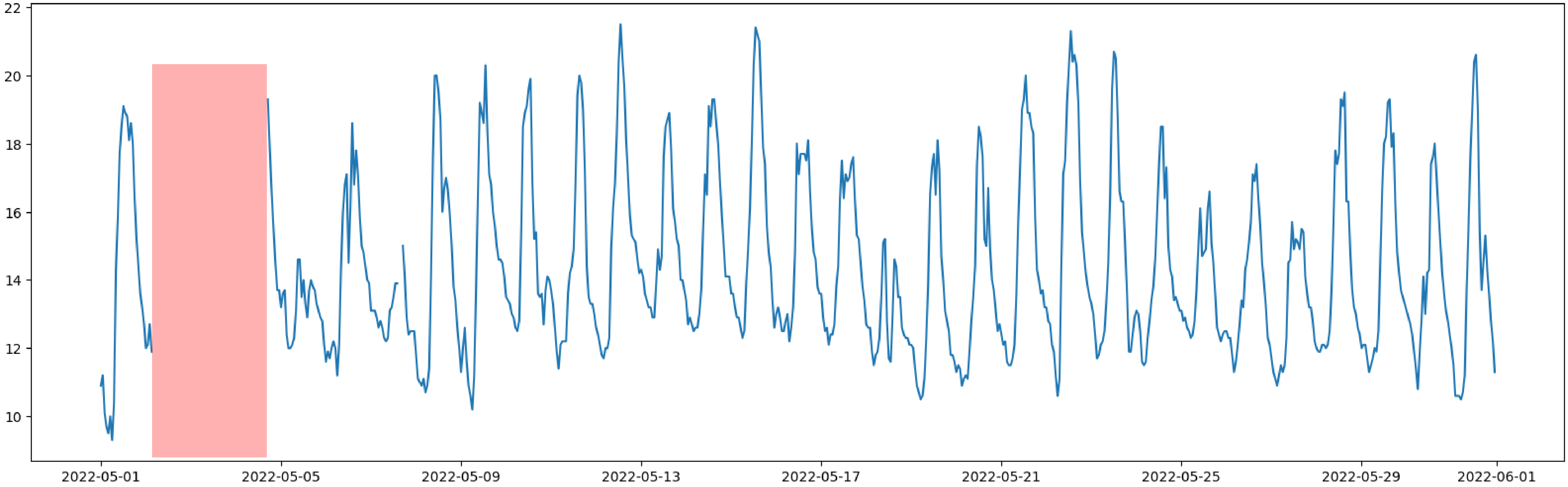
Se calcula usando una **media móvil** para suavizar los datos y eliminar fluctuaciones de corto plazo.

$$T_t = \frac{1}{k}\sum_{i=-m}^{m} Y_{t-i}$$



Completación de Datos en Series de Tiempo

La **completación de datos** en series temporales es el proceso de **llenar los valores faltantes (NaN)** en una serie cronológica, asegurando que la información tenga una continuidad temporal adecuada.
Esto es clave para poder aplicar modelos estadísticos o de aprendizaje automático, ya que muchos de ellos no toleran valores nulos.

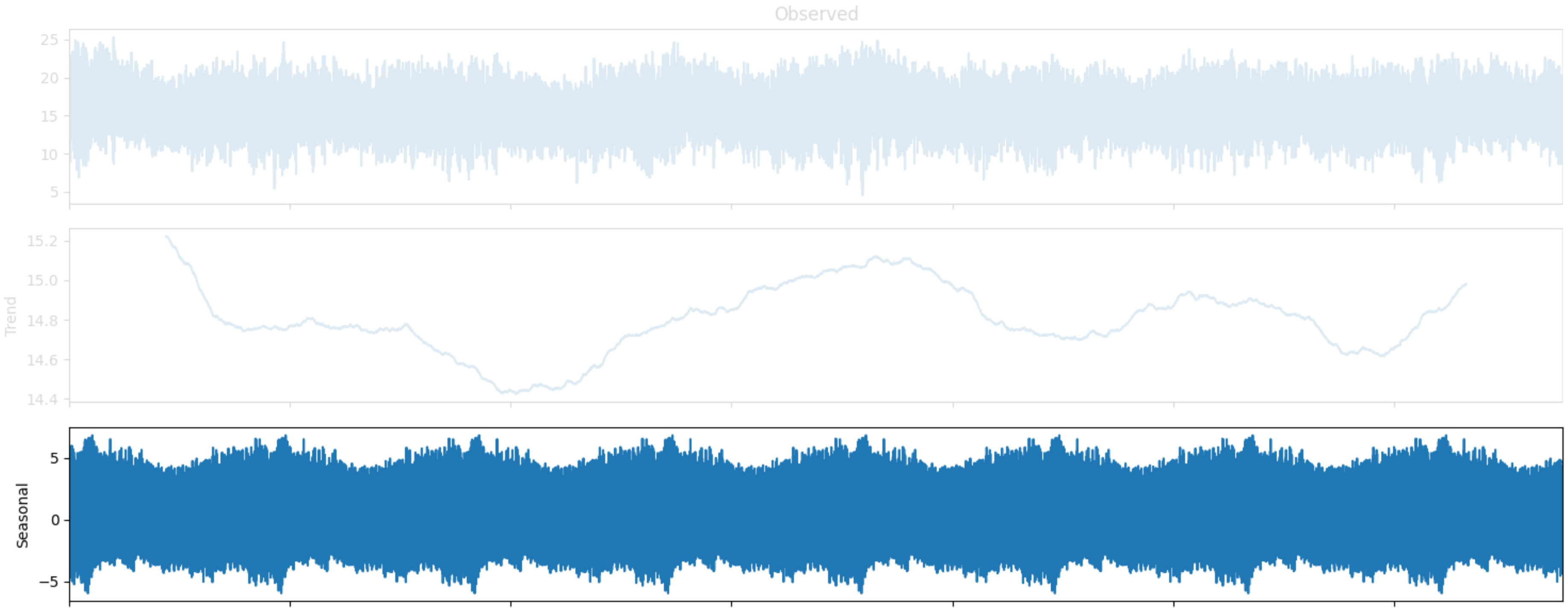


Modelos Estadísticos: Análisis de series temporales

Estacionalidad (S)

Son **patrones regulares y repetitivos** que se presentan en intervalos fijos, como días de la semana, meses del año o estaciones climáticas.

Se obtiene al **promediar los residuos de cada posición dentro del ciclo** (por ejemplo, los 12 meses si hay estacionalidad anual) **una vez eliminada la tendencia**.



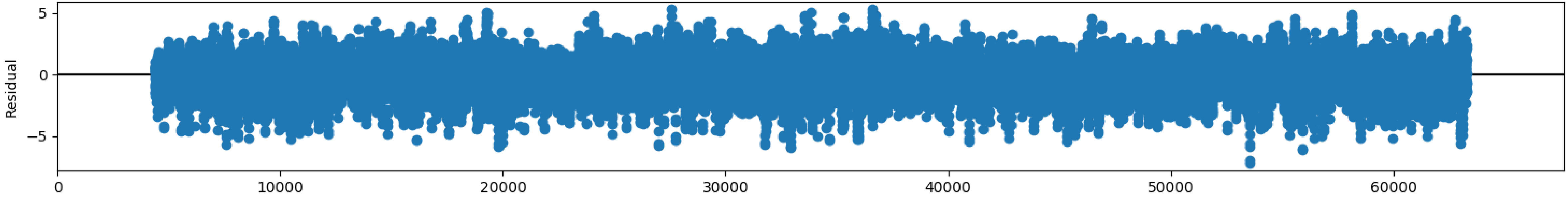
Ruido (R)

Es la **variabilidad aleatoria** o impredecible que no se explica por la tendencia ni por la estacionalidad. Se considera el "residuo" o error natural.

Una vez calculados los componentes de tendencia y estacionalidad, el residuo se obtiene como:

$$Rt = Yt - Tt - St$$

Este componente **debería comportarse como ruido blanco**, es decir, sin patrón ni autocorrelación.



Modelos Estadísticos: Análisis de series temporales

¿Para qué sirve el residuo (RUIDO) en el análisis de series temporales?

El residuo (o ruido) es lo que queda de la serie original después de eliminar la tendencia y la estacionalidad. Se define como:

$$Rt = Yt - Tt - St$$

¿Qué representa?

- El componente aleatorio o impredecible.
- Las perturbaciones externas no explicadas por patrones conocidos.
- Lo que un modelo no logra explicar.

¿Para qué se analiza el residuo?

1. Verificar si el modelo es adecuado:
   - Si el residuo no muestra patrón, el modelo (T + S) es bueno.
   - Si el residuo tiene estructura, puede que falte algo en el análisis.
2. Detección de eventos anómalos:
   - Valores extremos del residuo pueden indicar eventos inesperados (picos de contaminación, fallas, etc.).
3. Evaluar la aleatoriedad:
   - Un buen residuo debe parecer ruido blanco: sin autocorrelación y con media ≈ 0.

Modelos Estadísticos: Análisis de series temporales

¿Cuándo usar el análisis de series temporales?

El análisis de series temporales se usa cuando los datos están organizados cronológicamente y el orden en el tiempo afecta el comportamiento de la variable. Es decir, el tiempo no es una variable más, sino que la estructura temporal influye en la dinámica.

Se usa cuando:

- Se espera que los datos tengan patrones repetitivos (ciclos, estacionalidad).
- Hay tendencias a largo plazo que se desean analizar.
- Se quiere hacer predicción futura (forecasting).
- Es necesario detectar anomalías o cambios de comportamiento en el tiempo.

Campo	Variable temporal común
Meteorología	Temperatura, precipitación, viento, PM2.5
Energía	Demanda eléctrica, consumo de gas
Economía/finanzas	Precios, inflación, ventas
Salud pública	Casos diarios de enfermedad, hospitalizaciones
Medio ambiente	Concentración de contaminantes, descargas
Hidrología	Caudal diario, recarga hídrica, nivel de ríos