



ECO5008 Modelos predictivos

Clase 01 - Pronósticos y Simulación en la Gestión Hospitalaria

Sebastián Egaña Santibáñez 

Enlaces del profesor

-  <https://segana.netlify.app>
 -  <https://github.com/sebaegana>
 -  <https://www.linkedin.com/in/sebastian-egana-santibanez/>
-

Clase 01

Conceptos iniciales

Conceptos básicos de series de tiempo

Se habla de series de tiempo cuando nos referimos a una variable sobre la cual observamos su comportamiento a través del tiempo. Un ejemplo conocido, puede ser el principal indicador de productividad de un país, denominado Producto interno bruto (PIB). Veamos la siguiente imagen:

EL CICLO ECONÓMICO

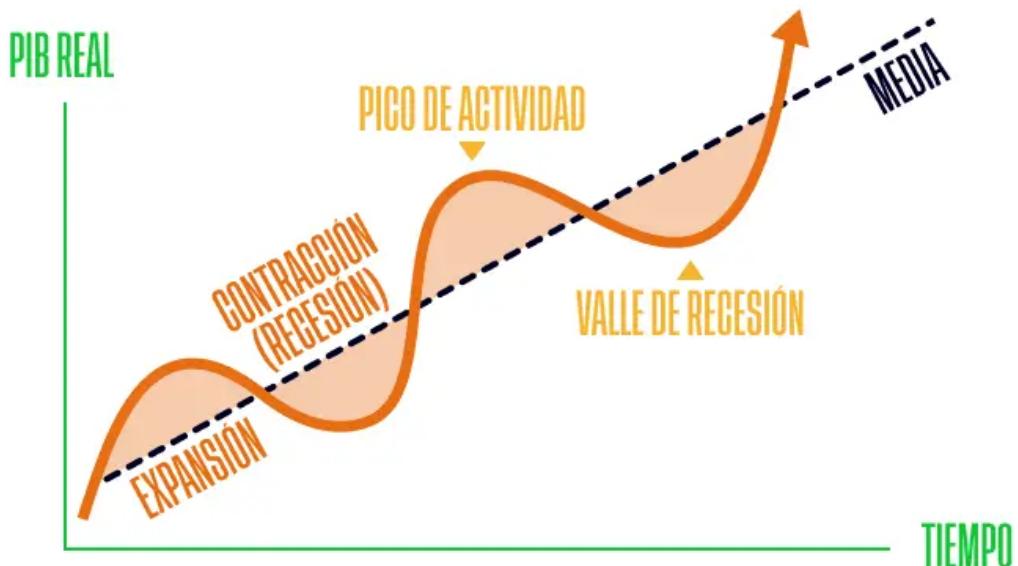


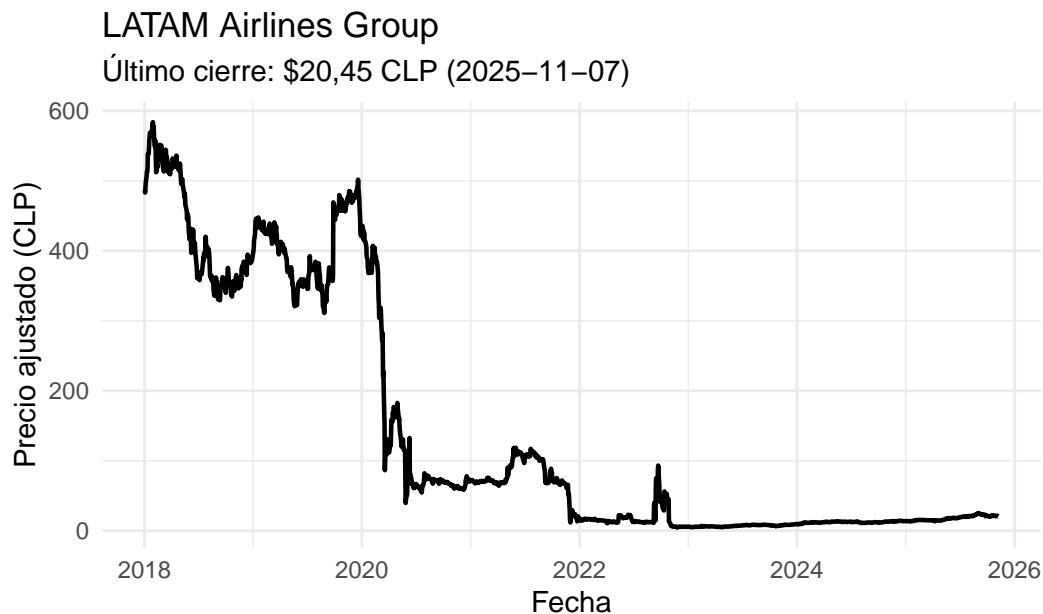
Figure 1: Obtenida desde: <https://revistapaideia.cl/ciclos-economicos-que-son-tipos-y-fases>

Pasemos ahora a hablar de ciertos componentes de las series de tiempo.

Tendencia, ciclo, estacionalidad y ruido

- Tendencia: corresponde al comportamiento general o a la dirección de largo plazo de la serie. Se debe tener cuidado que en algunas variables esto no siempre se cumple.

Veamos el caso para el precio de la acción de LATAM Airlines



Miremos ahora el caso en donde filtramos el período 2020 - 2021:



¿Qué pasa si filtramos desde el 2024 a la fecha?

LATAM Airlines Group

Último cierre en el rango: \$20,67 CLP (2025–09–25)



Fuente: Yahoo Finance vía tidyquant

Por esta razón debemos tener cuidado con apresurarnos a tomar conclusiones sobre una serie solo en base a la tendencia.

- Ciclo: corresponde a la variación propia de la serie en torno a la tendencia que no responde a frecuencias predecibles o fijas.
- Estacionalidad: corresponde a variaciones de la serie que poseen un comportamiento predecible ya sea semanal, mensual o anual.
- Ruido: corresponde al componente aleatorio e irregular que no puede ser explicado a través de los otros componentes que generalmente se relaciona con variaciones erráticas e impredecibles que no siguen un patrón.

¿Por qué analizamos una serie de tiempo?

Se debe justificar el por qué esto es necesario en el contexto de negocios. Por lo general, el interés científico sobre una serie puede darse por dos razones:

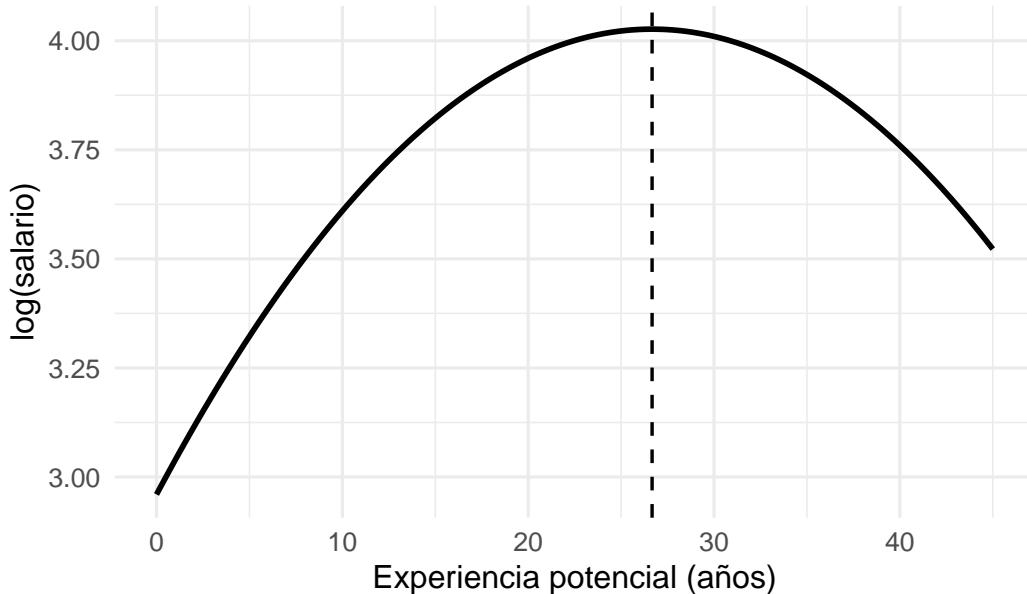
1. Explicar (por qué)
2. Predecir (qué y cuánto)

A pesar de esto, el límite entre ambas dinámicas es muy delgado. Veamos el siguiente ejemplo:

$$\ln w_i = \beta_0 + \rho S_i + \beta_1 X_i + \beta_2 X_i^2 + \gamma^\top Z_i + \varepsilon_i \quad (1)$$

- w_i : salario (normalmente por hora, en términos reales).
- S_i : años de escolaridad
- X_i : experiencia potencial (se utiliza también un coeficiente al cuadrado)
- Z_i : puede incluir variables de género, región, etc.

Curva minceriana: concavidad en log(salario)



Por lo general este modelo se utiliza para poder explicar factores que influyen dentro de la renta de las personas.

Pero ¿qué pasa si en vez de explicar la variable, asumimos que el modelo puede también predecir la variable de interés?

- Pensemos en el comportamiento tradicional que se hace entorno a los promedios de una variable.

Predecir, pronosticar y simular

Por lo general predecir se relaciona con contextos más generales que solo estimar el valor futuro de una serie temporal. Se puede referir a estimar el valor de una variable desconocida en distintos contextos temporales, como también estar asociado a variables no temporales (clasificación).

Para el caso de pronosticar, esto conecta por lo general con estimar el valor futuro de una variable (serie de tiempo). En este sentido pronóstico refiere a un contexto más acotado dentro de la dinámica predictiva. **En este sentido, todo pronóstico es una predicción, pero no toda predicción es un pronóstico.**

Una simulación refiere a la generación de múltiples valores futuros posibles a partir de supuestos probabilísticos para medir distribuciones, percentiles, etc.; con esto poder también determinar riesgo relacionado con los valores futuros.

Algunas técnicas de pronóstico:

Promedio móvil: suaviza datos, útil para detectar tendencia general.

$$\hat{y}_{t+1} = \frac{1}{k} \sum_{i=0}^{k-1} y_{t-i} \quad (2)$$

Suavizamiento exponencial simple: otorga más peso a datos recientes, capta mejor los cambios recientes

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t, \quad 0 < \alpha < 1 \quad (3)$$

Ejemplos numéricos

Datos y_t : 100, 104, 110, 120, 115, 118, 130, 128, 135, 140

Promedio

Pensemos en la situación de negocios más simple. Cálculamos el promedio y esto se convierte en nuestro pronóstico:

$$\hat{y}_{t+1} = 120$$

Promedio móvil ($k = 3$)

$$\hat{y}_{t+1} = \frac{1}{3}(y_t + y_{t-1} + y_{t-2})$$

- $\hat{y}_4 = \frac{110+104+100}{3} = 104.67$

- $\hat{y}_5 = \frac{120+110+104}{3} = 111.33$

- $\hat{y}_6 = \frac{115+120+110}{3} = 115.00$

Pronóstico $t = 11$:

$$\hat{y}_{11} = \frac{140 + 135 + 128}{3} = \mathbf{134.33}$$

Suavizamiento Exponencial Simple ($\alpha = 0.3$)

$$s_t = \alpha y_t + (1 - \alpha) s_{t-1}, \quad s_1 = y_1$$

- $s_1 = 100$
- $s_2 = 0.3 \cdot 104 + 0.7 \cdot 100 = 101.20$
- $s_3 = 0.3 \cdot 110 + 0.7 \cdot 101.20 = 103.84$
- $s_4 = 0.3 \cdot 120 + 0.7 \cdot 103.84 = 108.688$

Pronóstico $t = 11$:

$$s_{10} = \mathbf{129.628}$$

Actividad aplicada

Somos un equipo de analista y debemos que estimar una “buena manera” de poder saber cuántas atenciones tendrímos en el futuro por causas respiratorias. Utilizaremos el siguiente dataset del Gobierno de Chile:

- Atenciones de urgencias de causas respiratorias por semana epidemiológica

EDA (exploratorio)

Corresponde a **Exploratory Data Analysis**, lo que implica generar una primera impresión general de los datos.

Para esto podemos aplicar la siguiente librería **DataExplorer** [más info acá](#), considerando siempre primero importar el archivo.

```

library(readr)
library(tidyverse)
library(DataExplorer)
library(kableExtra)
library(knitr)
library(arrow)

#dataset_converted <- read_csv("data/dataset_converted.csv")
df <- read_parquet("data/at_urg_respiratorio_semanal.parquet")

df_filtered <- df %>%
  filter(Causa == "TOTAL CAUSAS SISTEMA RESPIRATORIO")

intro_table <- introduce(df_filtered)

intro_table %>%
  select(rows, columns, discrete_columns, continuous_columns,
         all_missing_columns, total_missing_values) %>%
  kable(
    format = "latex",
    booktabs = TRUE,
    caption = "Dataset summary from DataExplorer"
  ) %>%
  kable_styling(
    font_size = 5,
    latex_options = c("hold_position", "striped")
)

```

Table 1: Dataset summary from DataExplorer

| rows | columns | discrete_columns | continuous_columns | all_missing_columns | total_missing_values |
|--------|---------|------------------|--------------------|---------------------|----------------------|
| 317682 | 25 | 16 | 9 | 0 | 23700 |

El análisis más detallado se puede obtener con el siguiente código:

```
create_report(df_filtered)
```

Con esto podemos darnos una primera impresión de los datos que contiene el archivo descargado, como también algunas características de cada variable.

No centraremos en el total de atenciones para un centro de atención en particular; para esto generamos los siguientes filtro:

```

sotero_01 <-
  df_filtered %>%
  filter(RegionCodigo == 13
        & ServicioSaludCodigo == 14
        & ComunaCodigo == 13201
        & EstablecimientoGlosa == "Complejo Hospitalario Dr. Sótero del Río")
  select(EstablecimientoGlosa, Anio, NumTotal, Causa) %>%
  filter(Causa == "TOTAL CAUSAS SISTEMA RESPIRATORIO") %>%
  group_by(EstablecimientoGlosa, Anio) %>%
  summarise(total_casos = sum(NumTotal))

sotero_01 %>%
  kable(
    format = "latex",
    booktabs = TRUE,
    caption = "Atenciones por año"
  ) %>%
  kable_styling(
    font_size = 5,
    latex_options = c("hold_position", "striped")
  )
)

```

Table 2: Atenciones por año

| EstablecimientoGlosa | Anio | total_casos |
|--|------|-------------|
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2014 | 17671 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2015 | 18090 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2016 | 18479 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2017 | 19744 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2018 | 18586 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2019 | 22210 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2020 | 6324 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2021 | 5634 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2022 | 19852 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2023 | 17479 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2024 | 16332 |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2025 | 11258 |

```

graph_01 <-
  ggplot(sotero_01) +
  geom_line(aes(Anio, total_casos))

graph_01

```

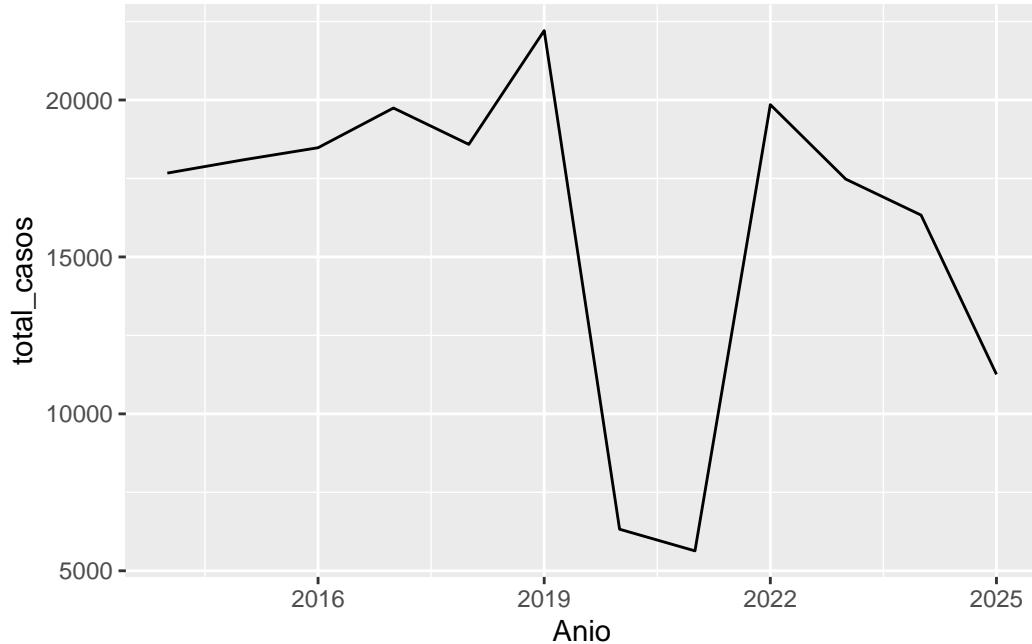


Table 3: Atenciones por año con pronósticos: PM ($k=3$) y SES ($\alpha=0.30$) para 2025

| EstablecimientoGlosa | Anio | Atenciones | PM(k) | SES(alpha) |
|--|------|------------|----------|------------|
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2014 | 17671 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2015 | 18090 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2016 | 18479 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2017 | 19744 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2018 | 18586 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2019 | 22210 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2020 | 6324 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2021 | 5634 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2022 | 19852 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2023 | 17479 | NA | NA |
| Complejo Hospitalario Dr. Sótero del Río (Santiago, Puente Alto) | 2024 | 16332 | NA | NA |
| NA | 2025 | NA | 17887.67 | 15825.07 |

Con esta simple metodología podemos plantear un pronóstico para lo que queda del año 2025.

Pregunta

1. ¿Qué conclusión puede sacar el promedio móvil? ¿Del suavizamiento?
2. ¿Cómo evaluaría el pronóstico generado?

Simulaciones de Monte Carlo

Una simulación corresponde a imitar en un entorno controlado para experimentar posibles comportamientos futuros de un sistema particular. En este caso lo que se simula es el valor futuro de la serie de tiempo que estamos intentando predecir.

El concepto de **Monte Carlo** corresponde al muestreo aleatorio repetido de múltiples valores futuro para poder incorporar incertidumbre dentro de nuestras predicciones.

En base a la serie ya analizada, definimos una simulación de montecarlo en base a lo siguiente:

1. Pronóstico base: calculas un número para el próximo año con dos métodos simples.
2. Error histórico: miras cómo se equivocaron esos métodos en el pasado (residuales = dato real – pronóstico de 1 paso). Esa lista de errores es tu “bolsa” de incertidumbre.
3. Simular futuros: repites miles de veces: sacas al azar un error de la bolsa (con reemplazo), se lo sumas al pronóstico base obtienes un posible futuro.
4. Resultado: miles de posibles valores futuros (una distribución).

Table 4: Pronóstico para 2025 con intervalos Monte Carlo (B=10000).

| Metodo | Pronóstico puntual | Mediana MC | P5 MC | P95 MC | Prob | capacidad |
|--------------|--------------------|------------|----------|----------|------|-----------|
| PM (k=3) | 17887.7 | 18286.67 | 8505 | 27136.33 | NA | |
| SES (=0.30) | 15825.1 | 16549.25 | 2506.169 | 23033.81 | NA | |

En términos de distribuciones:

Distribuciones predictivas por bootstrap

$B \sim 10,000$ simulaciones por método

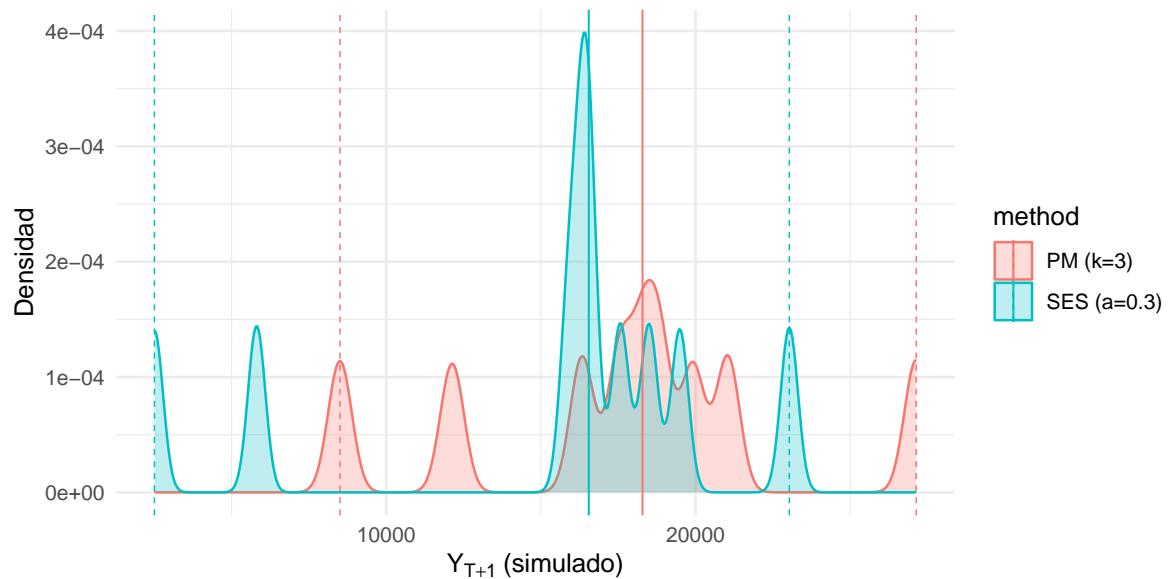


Figure 2: Distribuciones predictivas (bootstrap) para PM y SES: mediana (línea sólida) y P5/P95 (líneas punteadas).

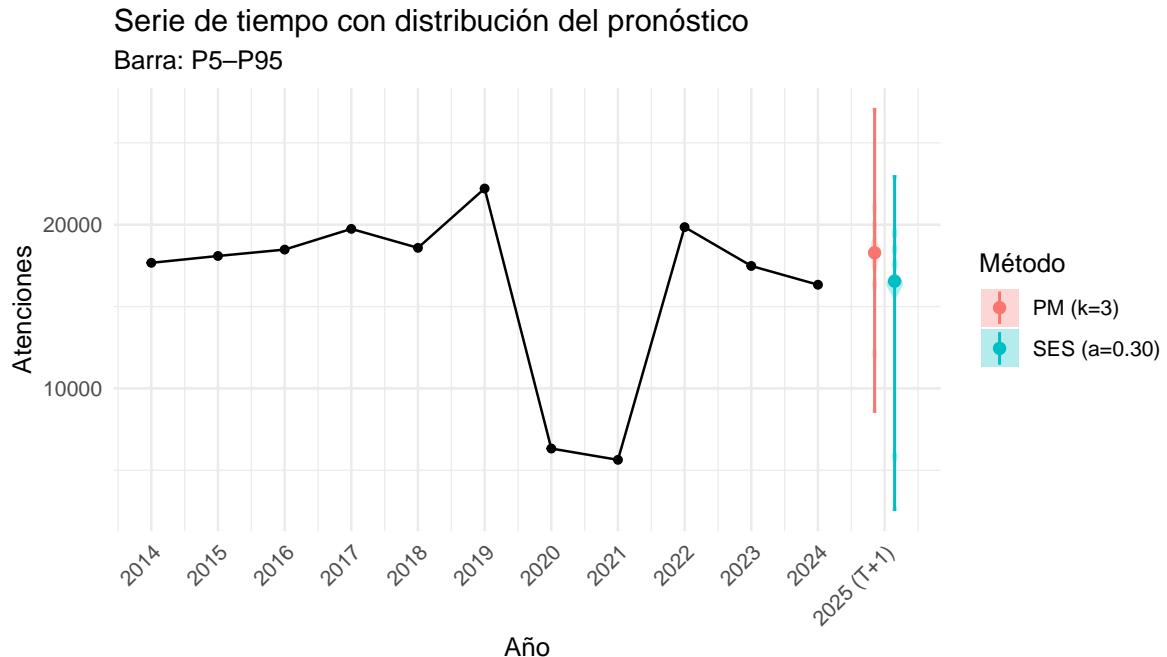


Figure 3: Serie anual con distribución del pronóstico en T+1 (PM y SES).

Para este caso, el principal valor de las simulaciones viene dado por la posibilidad de seleccionar métodos con un mayor o menor grado de incertidumbre; por otra parte, se podría también trabajar en sensibilizar el posible impacto de la predicción ya obtenida en base a los valores de relevantes de la simulación.

Referencias

- Vecillas Martin, D., Berrueto Fernández, C., & Gento Municio, A. M. (2025). Systematic Review of Discrete Event Simulation in Healthcare and Statistics Distributions. *Applied Sciences*, 15(4), 1861. [Enlace aquí](#)
- Heins, J., Schoenfelder, J., Heider, S., Heller, A. R., & Brunner, J. O. (2022). A scalable forecasting framework to predict COVID-19 hospital bed occupancy. *INFORMS Journal on Applied Analytics*, 52(6), 508-523. [Enlace aquí](#)
- Harper, P. R., & Shahani, A. K. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational research Society*, 53(1), 11-18.