



UNAM

Facultad de Ciencias

Temperatura Diaria en la Ciudad de Delhi 2013-2017

Alumno: Andrés González Ortega

Profesora: Sofía Villers Gómez

Ayudante: Carlos Fernando Vásquez Guerra

Ayudante: Midori Alondra Ortega Hernández

Este reporte fue elaborado como parte de la Tarea 3 en el curso
Análisis de Supervivencia y Series de Tiempo
en el Semestre 2025-2

23 de Marzo de 2024

Contents

1	Introducción	1
1.1	Contexto y Problemática	1
1.2	Objetivos	1
1.3	Metodología y Datos	2
2	Análisis y Modelos	3
2.1	Nuestro Proceso	3
2.2	Primer vistazo	3
2.3	Estacionalidad	5
2.4	Predicciones con nuestro modelo	5
2.5	Modelo con auto.arima	5
3	Conclusión y posibilidades	7
3.1	Conclusión	7
3.2	Trabajo Futuro	7

List of Figures

2.1	Gráfica principal de la serie de tiempo	3
2.2	Descomposición de la Serie	4
2.3	Autocorrelación en lag 60	4
2.4	Correlación Parcial lag 60	4
2.5	Predicciones	5
2.6	Errores del modelo	6
2.7	Predicciones ARIMA	6

Chapter 1

Introducción

El análisis de series de tiempo es una herramienta fundamental en la comprensión de fenómenos que varían a lo largo del tiempo. En este reporte, se presenta un estudio de la temperatura en la ciudad de Delhi durante el período de 2013 a 2017.

1.1 Contexto y Problemática

Delhi, conocida por sus extremos climáticos, experimenta veranos intensamente calurosos e inviernos fríos, con una marcada variabilidad estacional. Esta variabilidad hace que el análisis de la temperatura sea crucial para diversas aplicaciones, desde la planificación urbana hasta la salud pública.

Comprender las tendencias y patrones en la temperatura de Delhi no solo nos permite predecir futuros comportamientos climáticos, sino que también nos ayuda a identificar posibles anomalías y cambios a largo plazo. El análisis de series de tiempo proporciona una visión profunda de cómo las temperaturas han fluctuado, permitiendo a los ciudadanos y los responsables de políticas tomar decisiones informadas basadas en datos históricos.

1.2 Objetivos

El análisis de series de tiempo tiene como objetivo principal la comprensión y modelado de datos que varían a lo largo del tiempo. En el contexto del estudio de la temperatura en Delhi de 2013 a 2017, uno de los objetivos fundamentales es la predicción de posibles observaciones futuras. Esta capacidad predictiva es crucial para diversas aplicaciones prácticas y científicas.

- **Predicción de Tendencias y Patrones:** Al analizar la serie de tiempo de la temperatura, buscamos identificar tendencias y patrones recurrentes. Estos patrones pueden incluir ciclos estacionales, tendencias a largo plazo y variaciones anómalas. Comprender estos elementos nos permite anticipar cómo podría comportarse la temperatura en el futuro, basándonos en datos históricos.
- **Detección de Anomalías:** Otro objetivo importante es la detección de anomalías o eventos atípicos. Estos pueden ser indicativos de cambios climáticos significativos o de eventos extremos, como olas de calor o frío. La identificación temprana de estas anomalías permite a las autoridades y a la comunidad científica tomar medidas preventivas y de mitigación.
- **Optimización de Recursos:** La predicción precisa de la temperatura futura tiene beneficios inmediatos en la optimización de recursos. Por ejemplo, permite una mejor

planificación del uso de energía para calefacción y refrigeración, lo que puede resultar en ahorros significativos y en una gestión más eficiente de los recursos energéticos.

1.3 Metodología y Datos

Para llevar a cabo el análisis de la serie de tiempo, se utilizó el software R, aprovechando las potentes librerías `tseries` y `forecast`. Estas herramientas proporcionan un conjunto robusto de funciones y métodos para el análisis y modelado.

- **Análisis Preliminar:** El primer paso en nuestro análisis consistió en la exploración de los datos mediante diagramas de autocorrelación (ACF) y autocorrelación parcial (PACF). Estos diagramas nos permitieron identificar patrones y relaciones en los datos, así como determinar la presencia de estacionalidad y tendencias.
- **Modelado Inicial:** Para el modelado inicial, se emplearon las funciones `AM()` y `AR()` además de las transformaciones básicas (logaritmo, `diff` y `sqrt`), que nos ayudaron a ajustar modelos autorregresivos a los datos. Estos modelos proporcionaron una base para entender la estructura subyacente de la serie temporal.
- **Ajustes ARIMA:** Posteriormente, se realizaron ajustes utilizando modelos ARIMA (Autorregresivos Integrados de Media Móvil). Los modelos ARIMA son fundamentales en el análisis de series de tiempo, ya que combinan componentes autorregresivos (AR), de media móvil (MA) y de diferenciación (I) para capturar la dinámica de los datos.
- **Correlogramas y Residuales:** Se generaron correlogramas para evaluar la adecuación de los modelos ajustados. Estos gráficos mostraron las correlaciones residuales, permitiéndonos verificar si los modelos capturaban adecuadamente la estructura de los datos o si quedaban patrones sin modelar.
- **Predicción:** Finalmente, la herramienta fundamental de nuestro análisis fue la predicción. Utilizando los modelos ajustados, se realizaron predicciones de la temperatura futura. Estas predicciones proporcionan una visión detallada de los posibles comportamientos climáticos basados en los datos históricos.

Datos: El conjunto de datos utilizado para este análisis fue obtenido de [Kaggle](#) y contiene información climática diaria, 1462 observaciones de la temperatura media diaria registrada en la ciudad medida en grados Celsius, adicionalmente tenemos 144 observaciones más como conjunto de validación para nuestra predicciones y evaluar estas. Entre las diversas variables disponibles, se seleccionó la temperatura media, medida en grados Celsius, debido a su relevancia para el estudio de la estacionalidad en las series de tiempo. La temperatura presenta un ciclo claro acorde a las estaciones del año, lo que la hace ideal para este tipo de análisis.

Chapter 2

Análisis y Modelos

2.1 Nuestro Proceso

En este análisis, comenzaremos con un vistazo general a la serie de tiempo. Este paso inicial nos permitirá identificar patrones visibles y obtener una comprensión preliminar de los datos. A continuación, realizaremos una prueba de estacionalidad para determinar si la serie presenta componentes estacionales significativos. Si la prueba de estacionalidad no resulta concluyente, procederemos a analizar posibles transformaciones de los datos. Una vez que hayamos estabilizado la serie, evaluaremos los diagramas de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar la estructura subyacente de los datos. En base a ello se crearan modelos adecuados.

2.2 Primer vistazo

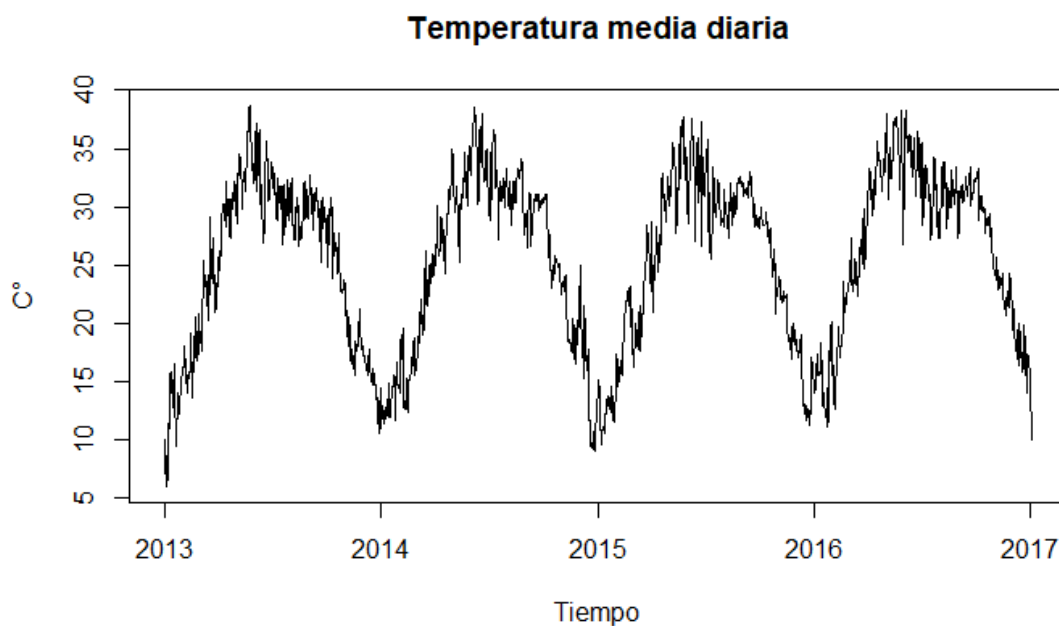


Figure 2.1: Gráfica principal de la serie de tiempo

Esta última gráfica presenta las 1462 observaciones diarias que se toman en grados Celsius desde el primer día de enero de 2013 al primer día de enero de 2017.

Ahora veamos la descomposición en tres partes, es decir, aislar la estacionalidad, la tendencia y el ruido blanco.

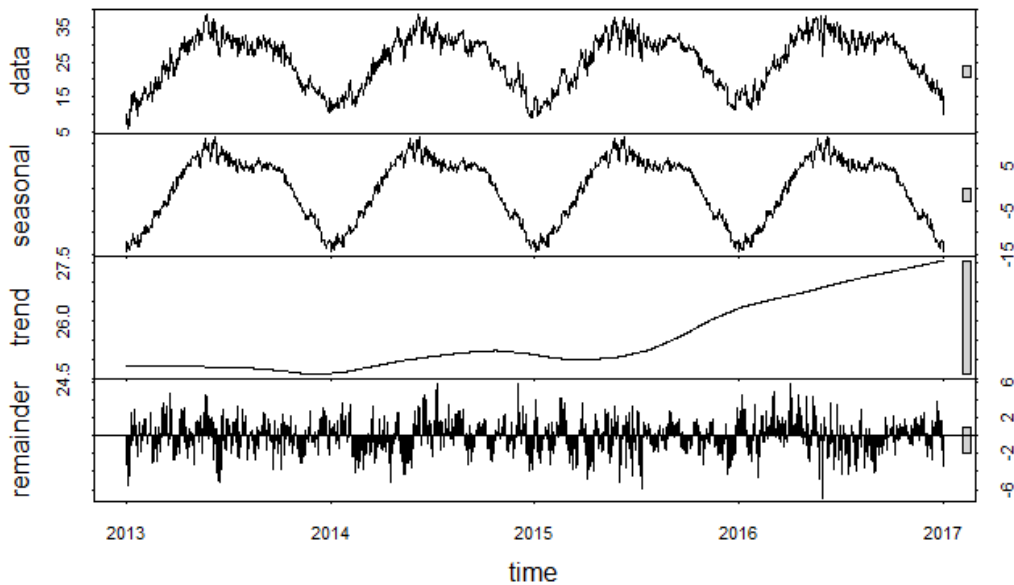


Figure 2.2: Descomposición de la Serie

Observamos que la estacionalidad de la serie es cíclica, correspondiendo a una periodicidad anual de las estaciones. En particular, las temperaturas son más bajas al final del año, es decir, durante el invierno. Otra observación importante es la tendencia al alza en los últimos dos años, lo que podría indicar un aumento extraño en la temperatura, una anomalía que habrá que tener en cuenta.

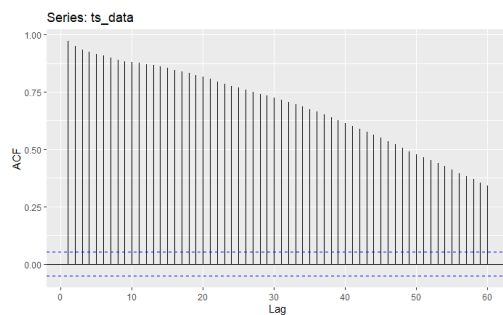


Figure 2.3:
Autocorrelación en lag 60

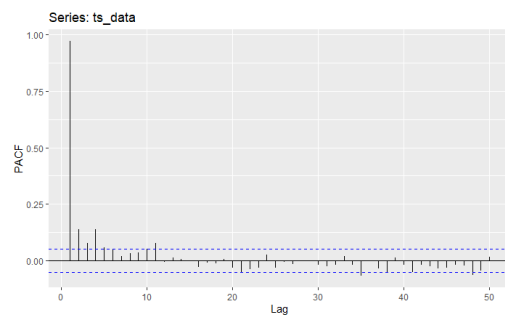


Figure 2.4:
Correlación Parcial lag 60

Así juntando la información de todos estos diagramas es viable proponer un modelo $AR(1)$, que tome en cuenta la última observación. El ACF presenta un desvanecimiento en la correlación parcial, ambos signos indican el modelo AR como el adecuado porque el corte en el PACF se realiza en el primer lag.

2.3 Estacionalidad

La prueba de estacionalidad es inconcluyente; se ha usado la prueba Dickey-Fuller, por lo que se ha realizado una transformación a la serie. La transformación ha sido la función diferenciación y con ello se consiguió la estacionalidad.

2.4 Predicciones con nuestro modelo

Se han usado los datos de entrenamiento y algunos intervalos de predicción para comparar.

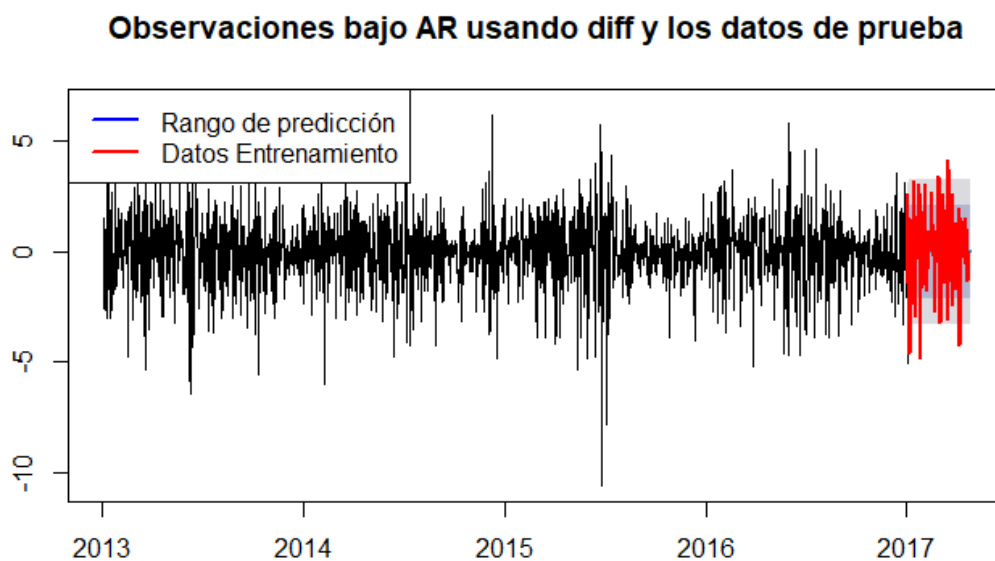


Figure 2.5: Predicciones

Las predicciones son favorables. Se alcanzan a apreciar algunos outliers de días donde la temperatura media se dispara para abajo o para arriba, podemos identificar estos outliers ya ocurridos como olas de calor y de frío. Con esta información, si en un futuro nuestras observaciones salen de ese rango $[-4,4]$, se tiene este fenómeno clarificado. Además, teniendo en cuenta nuestro modelo que solo se ajusta a los datos del primer día de 2017, este modelo para predicciones es satisfactorio. Sin embargo, podemos usar la función `auto.arima` para llegar a un modelo más adecuado.

2.5 Modelo con `auto.arima`

Existe una función en las paqueterías mencionadas que nos arroja un modelo preciso para modelar los datos y por lo tanto el más óptimo para realizar predicciones.

Después de aplicar la función nuestra serie de tiempo obtenemos el siguiente modelo $ARIMA(3,1,1)(0,1,0)[365]$. En la parte no estacionaria indica un 3 terminos autoregresivos, es decir tomamos en cuenta las ultimas tres observaciones para un nueva observación, el numero 1 después indica la diferenciación básica de 1 para conseguir la estacionalidad y el

valor para el termino de medias móviles es de 1. La parte estacionaria indica un diferenciación en estaciones de uno y el ciclo de nuestras estaciones es de 365 días o un año.

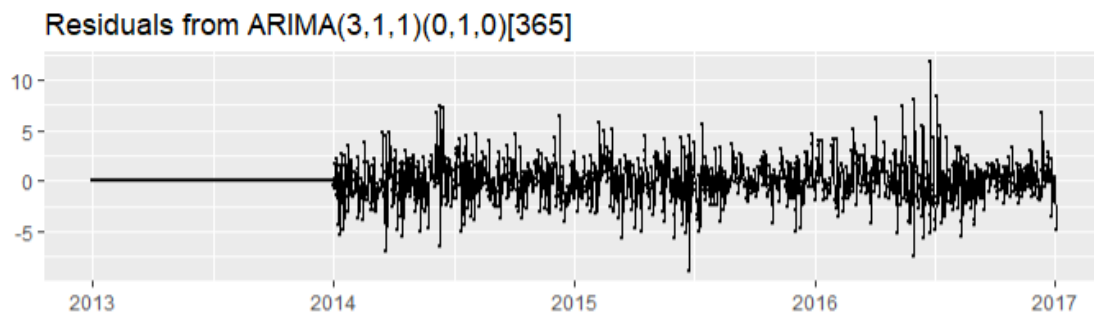


Figure 2.6: Errores del modelo

El análisis de las estadísticas del modelo y los residuales son favorables por lo que se procedió con las predicciones sin ningún margen de duda. Estos días con Residuales extremos nos abren camino para construir un modelo mejor o enfocarnos en las causas y relaciones que pudieron haber provocado tal anomalía en estudios futuros.

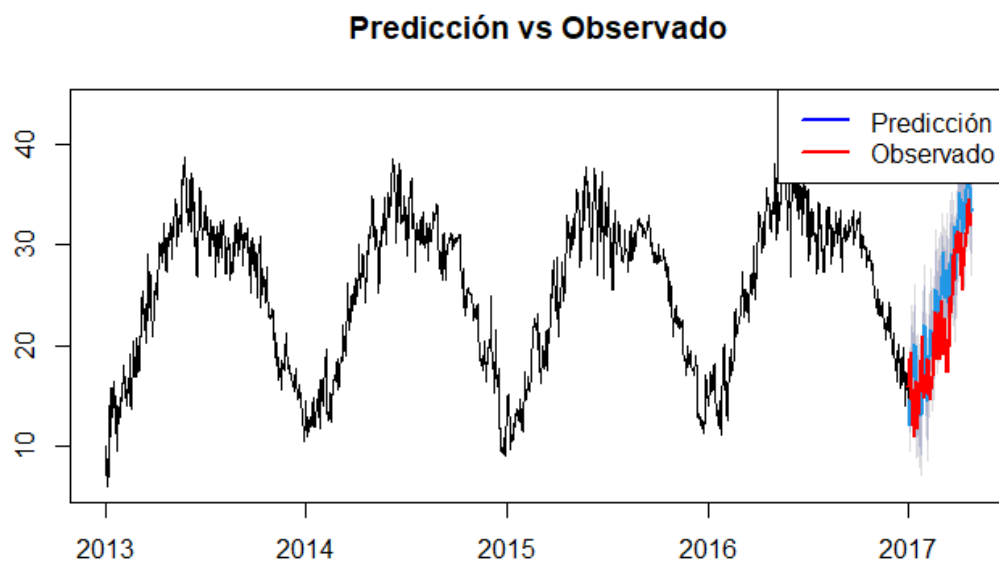


Figure 2.7: Predicciones ARIMA

Chapter 3

Conclusión y posibilidades

3.1 Conclusión

El modelo ARIMA que hemos desarrollado proporciona una predicción razonable de las temperaturas diarias, capturando tanto las tendencias estacionales como las fluctuaciones a corto plazo. A través del análisis de los correlogramas y la selección de los parámetros adecuados, hemos logrado ajustar un modelo que refleja fielmente el comportamiento histórico de la serie temporal.

Durante el proceso de modelado, identificamos varios outliers que corresponden a eventos extremos, como olas de calor y de frío. Estos outliers fueron detectados mediante el análisis de los residuos del modelo, donde observamos desviaciones significativas de los valores esperados. La identificación de estos eventos es crucial para mejorar la precisión del modelo y para entender mejor los patrones climáticos extremos.

En los últimos años, hemos observado una tendencia de crecimiento en las temperaturas. Esta tendencia es de suma importancia, ya que tiene implicaciones significativas para la planificación urbana, la agricultura y la salud pública. Comprender y predecir estas tendencias nos permite prepararnos mejor para los impactos futuros y desarrollar estrategias de mitigación adecuadas o para la prevención con nuestras predicciones.

3.2 Trabajo Futuro

Para profundizar en el análisis de las anomalías en la tendencia del crecimiento de la temperatura, proponemos extender este estudio a ciudades cercanas. Comparar los datos de diferentes localidades nos permitirá identificar patrones regionales y comprender mejor las causas subyacentes de las variaciones de temperatura. Además, es fundamental continuar recopilando más observaciones a lo largo del tiempo. Un conjunto de datos más extenso y actualizado mejorará la eficiencia del modelo y permitirá realizar predicciones más precisas y confiables.

Nuestro modelo ha demostrado ser una herramienta valiosa para la predicción de temperaturas diarias y la identificación de eventos climáticos extremos, así como una herramienta de predicción a un plazo corto.