

# Facultad de Ciencias

*Departamento de Estadística*

Carrera Ingeniería Estadística

## Proyecto de Título II

Asignatura	: Proyecto de Título II (220068)
Título	: Predicción de cantidad de llamadas en Contact Center usando métodos de Series de Tiempo
Realizado por	: Camila Molina Fernández Luisa Reyes Medina
Profesor Guía	: Sergio Contreras Espinoza
Profesor Co-Guía	: Francisco Novoa Muñoz
Semestre	: Segundo semestre 2018



UNIVERSIDAD DEL BÍO BÍO  
www.ubiobio.cl

# Índice

	<b>2</b>
<b>1. Introducción</b>	<b>5</b>
<b>2. Objetivos</b>	<b>7</b>
2.1. Objetivo General . . . . .	7
2.2. Objetivos Específicos . . . . .	7
<b>3. Antecedentes Generales</b>	<b>8</b>
<b>4. Definición Series de Tiempo</b>	<b>18</b>
<b>5. Definición MAPE</b>	<b>19</b>
<b>6. Resultados</b>	<b>20</b>
6.1. Suavizamiento exponencial . . . . .	20
6.1.1. Suavizamiento exponencial simple . . . . .	20
6.1.2. Suavizamiento exponencial doble . . . . .	22
6.2. Box-Jenkins . . . . .	26
6.2.1. Box-Jenkins, excluyendo últimas 5 observaciones . . . . .	27
6.2.2. Box-Jenkins, excluyendo últimas 10 observaciones . . . . .	34
6.3. Espacio Estado . . . . .	39
6.3.1. Espacio Estado, excluyendo últimas 5 observaciones . . . . .	39
6.3.2. Espacio Estado, excluyendo últimas 10 observaciones . . . . .	40
<b>7. Comparación</b>	<b>42</b>
<b>8. Conclusión</b>	<b>43</b>
<b>9. Códigos en R</b>	<b>44</b>

## Índice de figuras

1.	Cantidad de llamadas área comercial 2016 y 2017 . . . . .	9
2.	Cantidad de llamadas - Año 2017 . . . . .	10
3.	Cantidad de llamadas según día, por semanas del año . . . . .	12
4.	Cantidad de llamadas, días lunes por semanas del año . . . . .	12
5.	Cantidad de llamadas, días martes por semanas del año . . . . .	13
6.	Cantidad de llamadas, días miércoles por semanas del año . . . . .	13
7.	Cantidad de llamadas, días jueves por semanas del año . . . . .	14
8.	Cantidad de llamadas, días viernes por semanas del año . . . . .	14
9.	Cantidad de llamadas, días sábados por semanas del año . . . . .	15
10.	Cantidad de llamadas, días domingos por semanas del año . . . . .	15
11.	Cantidad de llamadas, Enero 2018 . . . . .	16
12.	Suavizamiento Exponencial Simple, excluyendo las últimas 5 observaciones . . . . .	21
13.	Suavizamiento Exponencial Simple, excluyendo las últimas 10 observaciones. . . . .	22
14.	Suavizamiento Exponencial Doble, excluyendo las últimas 5 observaciones. . . . .	24
15.	Suavizamiento Exponencial Doble, excluyendo las últimas 10 observaciones. . . . .	25
16.	Serie original, excluyendo las últimas 5 observaciones . . . . .	27
17.	Gráfico lag . . . . .	28
18.	ACF . . . . .	29
19.	Serie con logaritmo aplicado . . . . .	30
20.	Serie con primera diferenciación . . . . .	31
21.	Serie ajustada . . . . .	32
22.	Residuos . . . . .	33
23.	Serie con logaritmo aplicado . . . . .	35
24.	Serie ajustada, excluyendo las últimas 10 observaciones . . . . .	36
25.	Residuos . . . . .	37
26.	Gráfico lag . . . . .	38
27.	Muestra excluyendo últimas 5 observaciones, se observa el ajuste en rojo, predicción e intervalos de confianza del 90 % . . . . .	39
28.	Muestra excluyendo últimas 10 observaciones, se observa el ajuste en rojo, predicción e intervalos de confianza del 90 % . . . . .	40

## Índice de cuadros

1.	Cantidad de llamadas por horario en el año 2017 . . . . .	11
2.	Cantidad de llamadas por tipo de día - Año 2017 . . . . .	16
3.	Predicción sobre las últimas 5 observaciones . . . . .	21
4.	Predicción sobre las últimas 10 observaciones . . . . .	22
5.	Predicción sobre las últimas 5 observaciones . . . . .	24
6.	Predicción sobre las últimas 10 observaciones . . . . .	25
7.	Significancia de parámetros, excluyendo las últimas 5 observa- ciones . . . . .	31
9.	Modelo, excluyendo las últimas 5 observaciones . . . . .	32
8.	Comparación modelos ARIMA, excluyendo 5 observaciones . . .	32
10.	Predicción sobre las últimas 5 observaciones . . . . .	34
11.	Significancia de parámetros, excluyendo las últimas 10 observa- ciones . . . . .	35
12.	Comparación de modelos, excluyendo últimas 10 observaciones .	36
13.	Modelo, excluyendo las últimas 10 observaciones . . . . .	36
14.	Predicción, excluyendo las últimas 10 observaciones . . . . .	38
15.	Predicción sobre los últimas 5 observaciones . . . . .	40
16.	Predicción sobre las últimas 10 observaciones . . . . .	40
17.	Comparación MAPE, sobre últimas 5 y 10 observaciones . . . .	42
18.	Comparación MAPE, excluyendo últimas 5 y 10 observaciones .	42

# 1. Introducción

Nuestro problema surge de la necesidad de crear la distribución de turnos de los agentes de un Contact Center, para así asignar horarios de trabajos y los días de descanso de los trabajadores, una de las opciones que nos ofrece la estadística para este tipo de problemática, es predecir la cantidad de llamadas que se recibirán en el mes, así, al saber la cantidad de llamadas que se esperan, el encargado del área sabrá cuantos agentes necesita trabajando para cada día, y por consecuencia podrá designar los descansos y los turnos a los trabajadores.

Para poder realizar la predicción, utilizaremos métodos de Series de Tiempo, dentro de esta metodología existe la técnica de Suavizamiento Exponencial (Holt-Winters), Box-Jenkins y Espacio Estado, técnicas que aplicaremos a nuestros datos y que compararemos al final del estudio. Cabe destacar que la técnica a utilizar para la serie dependerá de las componentes que posea, una serie temporal puede poseer Nivel, Tendencia y Estacionalidad.

Según una clasificación realizada por Uriel Peiró sobre los métodos de predicción, las metodologías a utilizar en este estudio, están dentro de los métodos cuantitativos, dentro de la organización de métodos cuantitativos, existe la clasificación de análisis univariante, que es donde se encuentran clasificados los métodos que utilizaremos en este estudio, los cuales son: método de suavizamiento exponencial, donde existen tres formulaciones, suavizamiento exponencial simple, suavizamiento exponencial doble y suavizamiento exponencial triple, también aplicaremos el método de Box y Jenkins y el método de Espacio Estado.

Para llevar a cabo el análisis se considerará la serie de tiempo conformada por el número de llamadas recibidas por el Contact Center de manera diaria dentro del horario hábil (ahora desde 08:30 hrs. a 22:00 hrs.), luego y de ser necesario, separaremos los días que componen la serie según su naturaleza, debido a que existen irregularidades que alteran de manera significativa nuestra serie.

Para decidir de mejor manera el desarrollo y formulación del estudio, además de elegir el método a utilizar, es clave la información que se obtenga de la base de datos y de los gráficos, por lo que en primera instancia se procederá a describir y graficar las series obtenidas de la serie original y luego se repetirá el procedimiento para las series resultantes de la separación de datos, teniendo esto, se desarrollará un ajuste con cada uno de los métodos ya mencionados, estimando los parámetros que sean necesarios para cada ajuste, posterior a la ejecución de los ajustes, se calcularán los errores de cada método y también se efectuarán las predicciones, con el fin de comparar y elegir el método con menor error o con mejor predicción.

Cabe destacar que el Contact Center efectuó un cambio en el horario de atención en el año 2017, ampliando este, por lo que no se podrán utilizar la base de datos de años anteriores al 2017, ya que al ser más acotado el horario de atención al cliente, se reciben menos llamadas y esto impacta de manera negativa a nuestro ajuste y predicción, teniendo esto en cuenta, utilizaremos la base de datos conformada desde el 01 de Enero de 2017 hasta el 20 de Abril de 2018.

En resumen, se tiene la base de datos de la cantidad de llamadas en los días que clasificamos como “normales”, la cual consta de 151 observaciones, se procederá a aplicar los métodos ya mencionados, separando de la serie original las últimas 5 y 10 observaciones, con el fin de comparar los ajustes y calcular los *MAPE* respectivos en cada ajuste y así, seleccionar el método que mejor se ajuste a nuestros datos.

## **2. Objetivos**

### **2.1. Objetivo General**

Predecir la demanda en la central de atención telefónica de la empresa usando modelos de series de tiempo.

### **2.2. Objetivos Específicos**

- Constatar si es factible utilizar modelos de series de tiempo para el pronóstico de demanda de llamadas en la central de atención telefónica del área comercial de la empresa, considerando la condición de los días en que se reciben las llamadas.
- Diseñar pronósticos de demanda mediante el método de Suavizamiento Exponencial, la metodología de Box-Jenkins (ARIMA) y el método Espacio Estado, para luego comparar sus resultados y decidir que procedimiento ofrece el mejor resultado de predicción.
- Determinar que método de series de tiempo genera menores errores para predecir la serie estudiada.

### 3. Antecedentes Generales

La base de datos a utilizar es procedente de una compañía del área de telecomunicaciones de origen regional, la cual brinda servicios de Internet, Telefonía y Televisión digital a familias y empresas en más de 25 localidades de la Octava Región y sus alrededores. Cada uno de los servicios brindados por la empresa son cada vez más importantes en la vida diaria de las personas, de las organizaciones y de los países. Las telecomunicaciones fomentan el desarrollo social y económico, mejoran los servicios de salud, de educación, de gobierno y la democracia, estimulan la creación de empleos, contribuyen a preservar el medio ambiente, entre muchos otros beneficios. Por eso, no es posible imaginarse el presente o el futuro sin telecomunicaciones.

Si consideramos la información que nos entrega la página web de la compañía, podemos afirmar que actualmente es la única compañía que ha llegado a los sectores rurales más alejados, que es precisamente el objetivo impuesto, además de la variedad de sus servicios, obteniendo de esta manera, cada vez más personas que los prefieren. Cabe mencionar que la compañía cuenta con el área de servicio al cliente, el cual existe en cuatro modalidades, la modalidad presencial, la página web de la empresa, redes sociales y el Contact Center, siendo este último el más utilizado por los clientes, ya que es más fácil y rápido para obtener respuestas.

A pesar de la fuerte competencia que existe en el rubro de las telecomunicaciones, esta compañía está en constante crecimiento y además está en pleno desarrollo, por lo cual existen constantes cambios en las ofertas de sus servicios, como así también en la cantidad de llamadas recibidas en el Contact Center.

Con el fin de ayudar y facilitar la distribución de turnos de los agentes del Contact Center y a su vez asignar los horarios de trabajo y los días de descanso de los trabajadores, una alternativa de solución para este problema es predecir la cantidad de llamadas que se recibirán en el mes, así, al saber la cantidad de llamadas que se esperan, el encargado del área sabrá cuantos agentes necesita trabajando para cada día, y por consecuencia podrá designar los descansos y los turnos a los trabajadores.

Para poder realizar la predicción que planteamos, utilizaremos tres métodos de Series de Tiempo, los cuales son Suavizamiento Exponencial (Holt-Winters), Box-Jenkins y Espacio Estado, técnicas que aplicaremos a nuestros datos y que compararemos al final del estudio para elegir la que mejor se ajusta y la que mejores predicciones realiza.



En un principio, se tenía la base de datos del área comercial tanto como del área técnica de la compañía, se nos pidió trabajar con la base de datos del área comercial que era la prioridad en ese momento, se analizó la base de datos de llamadas del área comercial conformada desde el 1 de enero de 2016 hasta el 15 de noviembre de 2017, dentro del horario hábil u horario de trabajo del Contact Center, con observaciones mensuales, mostrando la Figura 1.

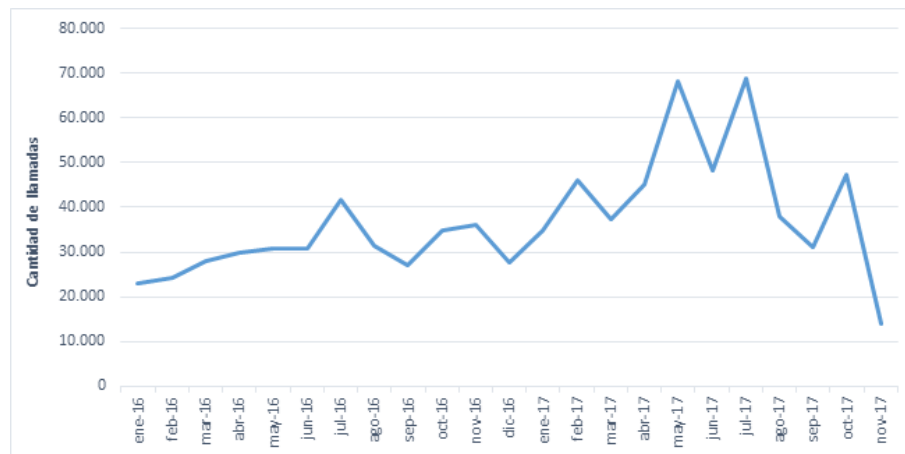


Figura 1: Cantidad de llamadas área comercial 2016 y 2017

Como muestra la Figura 1, se aprecia un notable aumento en la cantidad de llamadas entrantes en horario hábil, este aumento se debe al cambio de horario de atención del contact center (ahora de 08.30 a 22.00 hrs de lunes a domingo), lo que a la vez vuelve irrelevantes los datos del año 2016, ya que con este cambio se espera considerables aumentos de llamadas, por ende utilizaremos los datos a partir del año 2017 para realizar las predicciones.

Luego de completar la base de datos de llamadas del año 2017 hasta el 31 de Diciembre del mismo (ya que nuestra base de datos, estaba conformada hasta el 15 de noviembre de 2017), se graficaron los 365 días del año con sus respectivas frecuencias de llamadas. Obteniendo la Figura 2.

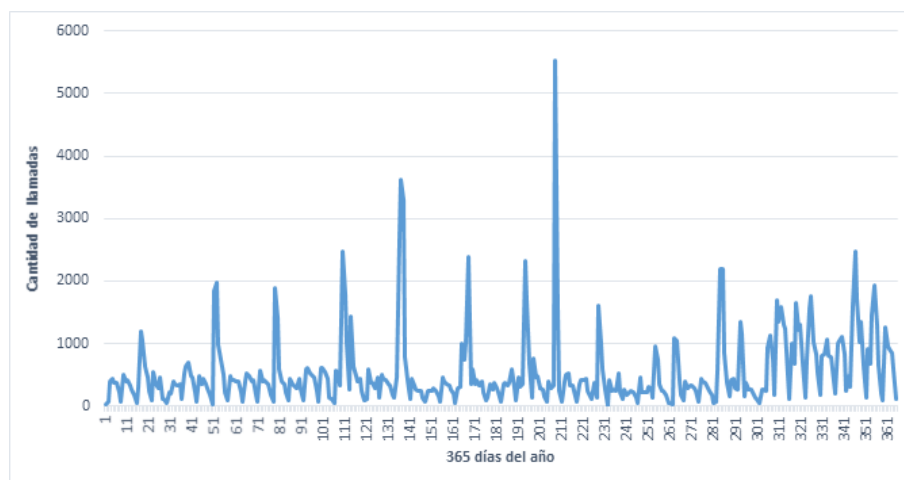


Figura 2: Cantidad de llamadas - Año 2017

La Figura 2 muestra los 365 días del año enumerados, mientras en el eje y se muestran las frecuencias de llamadas de los 365 días del año.

Consideramos contingencia como algún suceso imposible de controlar por parte de la compañía que afecte en la prestación de sus servicios, por ejemplo, lluvias torrenciales, cortes de electricidad, cortes de fibra óptica, etc. El punto atípico más alto mostrado en la Figura 2, corresponde a una contingencia el día 27 de Julio de 2017 que afectó a varios nodos, los puntos anteriores con mayor frecuencia corresponden a contingencias y/o días de suspensión de servicios días en los cuales tiende a ser común recibir más llamadas de lo habitual, pero que en el caso de las contingencias escapa del control de la compañía.

Se observa el notable aumento en las observaciones en el último período del año, este cambio es explicado por cambio de horario de atención del Contact Center, ya antes mencionado.

Luego de solamente considerar las llamadas del año 2017 en adelante, a continuación en el Cuadro 1 se detalla la cantidad de llamadas por intervalo de hora durante todo el año:

<b>Horario</b>	<b>Cantidad de llamadas</b>
8:30:00	1.473
9:00:00	7.951
10:00:00	11.666
11:00:00	14.473
12:00:00	15.523
13:00:00	14.099
14:00:00	13.664
15:00:00	14.044
16:00:00	13.654
17:00:00	16.247
18:00:00	14.246
19:00:00	11.740
20:00:00	8.913
21:00:00	6.792
22:00:00	3.683
23:00:01 a 08:29:59	117
<b>Total</b>	<b>168.266</b>

Cuadro 1: Cantidad de llamadas por horario en el año 2017

Con un total de 168.266 llamadas recibidas en todo el año 2017, las horas punta fueron dentro del intervalo de las 12 : 00 : 00 a 12 : 59 : 59 hrs. con 15.523 llamadas y las 17 : 00 : 00 a 17 : 59 : 59 hrs. con 16.247 llamadas.

Para poder llevar a cabo la predicción de Cantidad de llamadas, fue necesario separar la base de datos detalladamente por semanas (4 semanas de cada mes) con sus respectivas frecuencias o cantidad de llamadas (en total 52 semanas en el año).

Se separó de manera tal que se pudo comparar el comportamiento según el día semana a semana (52 semanas anuales, en el caso de los domingos son 53), mostrando que el día jueves es el que se destaca a lo largo del año, como muestra la Figura 3.

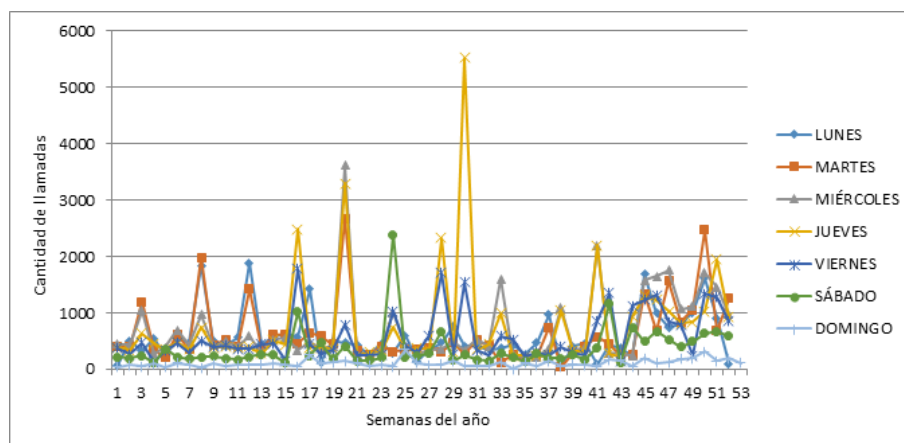


Figura 3: Cantidad de llamadas según día, por semanas del año

Para poder analizar individual y detalladamente la cantidad de llamadas según el día de la semana, en las siguientes figuras se muestran la cantidad de llamadas según días de la semana a lo largo del año 2017, comenzando con la Figura 4 que muestra la cantidad de llamadas de los días Lunes del año 2017.



Figura 4: Cantidad de llamadas, días lunes por semanas del año

La Figura 4 muestra la frecuencia de llamadas de los 52 días lunes del año 2017, se pueden observar varios puntos atípicos explicados por Contingencias o días de suspensión de Clientes. El gráfico muestra un aumento de llamadas en la semana 50, el cual corresponde al Lunes 11 de Diciembre, el cual fue día de contingencia que afectó a varios nodos.



Figura 5: Cantidad de llamadas, días martes por semanas del año

La Figura 5 muestra la cantidad de llamadas de todos los días martes del año. Podemos ver en un principio bastantes puntos atípicos, luego la serie varía dentro de los valores frecuentes, al final igual que en los días lunes se ve un aumento en la cantidad de llamadas entrantes al contact center, que como ya es sabido, se deben al cambio de horario de atención del mismo. Se observa que en la semana 20 (16 de Mayo) se registraron 2670 llamados, que corresponde a día de suspensión de clientes. El aumento a la alta cantidad de llamadas recibidas en las últimas 4 semanas del año, se debe a que hubo contingencia en 3 de estas 4 semanas.



Figura 6: Cantidad de llamadas, días miércoles por semanas del año

La Figura 6 nos muestra la frecuencia de llamadas de los días Miércoles, con puntos atípicos que son bastante variables, variabilidad que deriva de la envergadura o impacto de la contingencia (a la cantidad de Nodos que afecta), al igual que los días anteriores se ve un aumento en la cantidad de llamadas las últimas semanas del año. Se observa que en la semana 20 (17 de Mayo) se registraron 3.617 llamadas, debido a ser día post-suspensión de clientes.

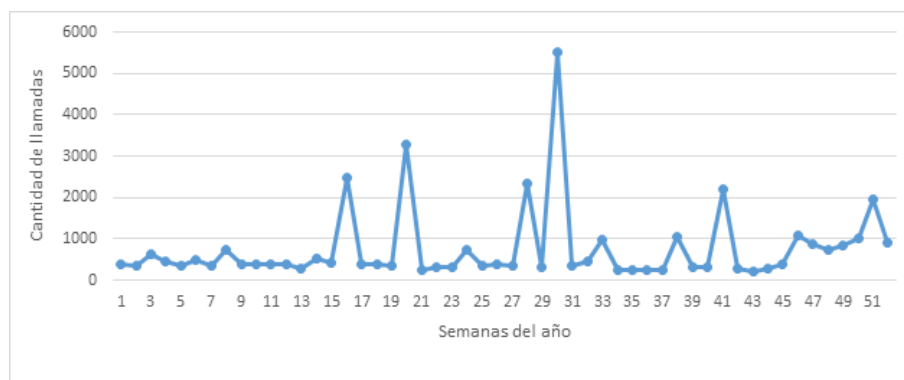


Figura 7: Cantidad de llamadas, días jueves por semanas del año

La Figura 7 muestra la cantidad de llamadas los días Jueves durante todo el año 2017, como en los casos anteriores, los puntos atípicos se deben a contingencias y a días de suspensión de clientes, al término de la serie se observa el mismo fenómeno que en las series anteriores, el aumento en la cantidad de llamadas entrantes. En la semana 30 (27 de Julio) se registraron 5.529 llamadas, debido a una contingencia que afectó a gran parte de los clientes, también se debe considerar, que existen días de suspensión de clientes y post suspensión de clientes, como a su vez contingencias de distintas magnitudes.



Figura 8: Cantidad de llamadas, días viernes por semanas del año

La Figura 8 nos muestra gráficamente la cantidad de llamadas de los días viernes del año, el comportamiento de estos datos es bastante distinto respecto a los gráficos o figuras anteriores, aunque se repite el aumento en llamados al final del año. Se observa bastante variabilidad en el comportamiento del día viernes, considerando también, que existen días post suspensión de clientes, como a su vez contingencias de distintas magnitudes.

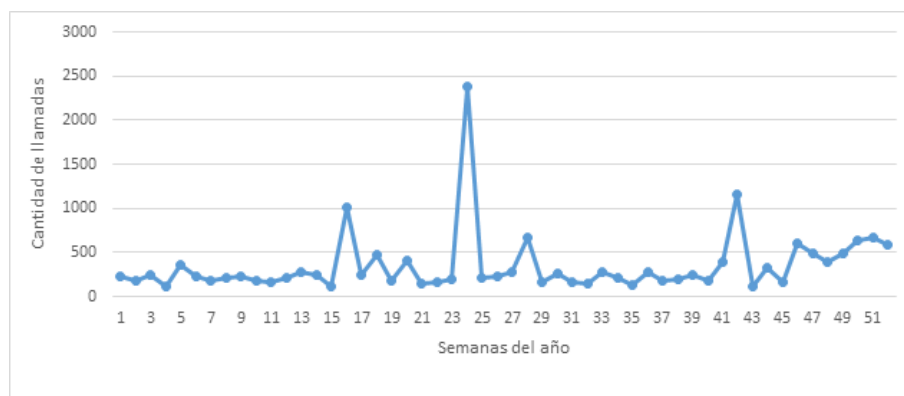


Figura 9: Cantidad de llamadas, días sábados por semanas del año

La Figura 9 muestra la cantidad de llamadas recibidas los días sábados, aunque en comparación la cantidad de llamadas de los días sábados versus los gráficos ya estudiados es notablemente inferior, se observa comportamiento similar en las últimas semanas del año. Los puntos más destacados corresponden a contingencias de distintas magnitudes, lo que explica el por qué la variabilidad en el comportamiento del día sábado.



Figura 10: Cantidad de llamadas, días domingos por semanas del año

Por último la Figura 10 muestra la cantidad de llamadas atendidas el día Domingo, la cantidad de llamadas recibidas los días domingos es bastante menor que la cantidad de llamadas del día sábado, pero al igual que los días anteriores, se repite el patrón de aumento en la cantidad de llamadas al término del año. A pesar de que normalmente el día domingo se reciben mucho menos llamados, también tienen días altos debido al efecto de las contingencias.

A continuación, mostraremos el Cuadro 2 que muestra el resumen de cantidad de llamadas por día de la semana del año 2017:

<b>Tipo de día</b>	<b>Cantidad de llamadas</b>
Días normales	61.274
Sábado	6.986
Domingo	3.054
Feriado	2.195
Suspensión de clientes	19.288
Día post suspensión	15.381
Contingencia	24.164
Día post contingencia	35.924
<b>Total</b>	<b>168.266</b>

Cuadro 2: Cantidad de llamadas por tipo de día - Año 2017

Luego de ver el análisis detallado de los días de la semana, procederemos a analizar gráficamente el mes de enero de 2018, como muestra la Figura 11.

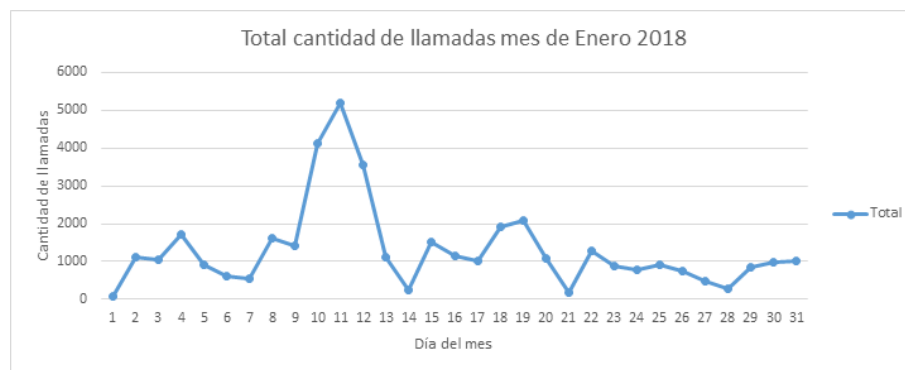


Figura 11: Cantidad de llamadas, Enero 2018

En el caso del día 1 de Enero se debe a que fue día feriado (a pesar de que el contact center atendió en horario normal), así como el día 11 de enero que fue día post contingencia, los días con menor cantidad de llamadas son los días 7,14,21 y 28 que son días domingo, lo que explica la baja frecuencia de llamadas, tal como analizamos en el gráfico anterior, el día domingo es el día en el que menos llamadas se reciben.



Al Analizar la cantidad de llamadas día a día según la condición de estos y al percatarnos que afectaban notablemente las series, fue necesario separar los días según su naturaleza, se muestra la definición de la naturaleza de los días de la semana a continuación:

- **Normales:** días de semana hábiles (lunes a viernes), sin contingencias ni suspensiones.
- **Feriados:** días feriados y festivos.
- **Contingencia:** días de contingencia o incidencia (imposibles de predecir).
- **Post contingencia:** considerando un comportamiento anormal de la cantidad de llamadas al día siguiente de contingencia (dependiendo el impacto de este), fue necesario considerar días de estas naturaleza aparte al resto.
- **Suspensión:** días de suspensión de clientes (considerable aumento en llamadas recibidas).
- **Post suspensión:** día siguiente a la suspensión de clientes.
- **Sábado:** días sábados hábiles, sin contingencia, existe la necesidad de separar los días sábados del resto de la semana debido a la baja cantidad de llamadas registradas.
- **Domingo:** días domingos, exceptuando los domingos que coincidn con un festivo, feriado o contingencia.

Considerando que los días de contingencia son impredecibles (ya que escapan del control de la compañía), en consecuencia los días post contingencias también se vuelven impredecibles, por ende los consideramos irrelevantes al momento de realizar predicciones.

En el caso de los días de suspensión de clientes y los días post-suspensión de clientes, se pueden programar, además en esos días siempre se reciben bastantes más llamadas de lo común, por ende, está demás crear predicción para esos días, ya que se sabe de antemano que las llamadas son muchas.

En el caso de los días Feriados, Sábados y Domingos, se tienden a recibir pocas llamadas (como ya se analizó gráficamente), por ende no es prioritario hacer predicciones para esos días.

Por último consideramos que los días normales son más volátiles en lo que a cantidad de llamadas recibidas respecta, además son los días con mayor presencia en nuestra base de datos, por ende, para este estudio haremos predicciones sólo para los días normales, ya que los consideramos más relevantes que el resto de los días ya definidos.

## 4. Definición Series de Tiempo

A continuación se presentarán dos definiciones de series de tiempo:

Una serie de tiempo es una secuencia de observaciones tomadas secuencialmente en el tiempo. Muchos conjuntos de datos aparecen como series de tiempo: una secuencia mensual de la cantidad de mercancías enviadas desde una fábrica, una serie semanal de la cantidad de accidentes de tráfico, cantidades diarias de lluvias, observaciones por hora sobre el rendimiento de un proceso químico, etc. Los ejemplos de series temporales abundan en campos como la economía, los negocios, la ingeniería, las ciencias naturales (especialmente la geofísica y la meteorología) y las ciencias sociales. Una característica intrínseca de una serie de tiempo es que, típicamente, las observaciones adyacentes son dependientes. La naturaleza de esta dependencia entre las observaciones de una serie temporal es de considerable interés práctico. El análisis de series de tiempo se refiere a técnicas para el análisis de esta dependencia. Esto requiere el desarrollo de modelos estocásticos y dinámicos para datos de series de tiempo y el uso de dichos modelos en áreas importantes de aplicación.<sup>1</sup>

Una serie de tiempo es un conjunto de observaciones  $y_1, \dots, y_n$  ordenado en tiempo. El modelo básico que representa una serie de tiempo es el modelo aditivo.<sup>2</sup>

$$y_t = \mu_t + \gamma_t + \epsilon_t, t = 1, \dots, n. \quad (1)$$

Donde:

- $\mu_t$  es un componente que varía lentamente llamado tendencia.
- $\gamma_t$  es una componente periodica de período fijo llamada estacional.
- $\epsilon_t$  es una componente irregular llamada error o perturbación.

---

<sup>1</sup>*Time series analysis: forecasting and control. George E.P.Box, Gwilym M.Jenkins, Gregory C.Reinsel, Greta M.Ljung*

<sup>2</sup>*Time series analysis by State Space Methods. J. Durbin and S.J. Koopman*

## 5. Definición MAPE

MAPE (Mean Absolute Percentage Error)

El error absoluto porcentual medio, nos ayudará a elegir el mejor modelo, buscando siempre que este indicador sea el menor posible, ósea, que el modelo cometa el menor porcentaje de error posible. El MAPE se define de la siguiente forma:

$$\frac{\sum_{i=1}^n \left| \frac{Y_t - F_t}{Y_t} \right| * 100}{n} \quad (2)$$

Donde:

- $Y_t$ : observación en el instante  $t$ .
- $F_t$ : predicción en el instante  $t$ .
- $n$ : cantidad de observaciones.

## 6. Resultados

### 6.1. Suavizamiento exponencial

#### 6.1.1. Suavizamiento exponencial simple

Este método se usa en caso de que la serie no presente tendencia ni estacionalidad, está definido de la siguiente manera:

$$F_t = \alpha * Y_t + (1 - \alpha) * F_t \quad (3)$$

Donde  $Y_t$  es la observación en el instante  $t$  y  $F_t$  es la predicción para la observación del instante  $t$ .

Con la condición de inicialización

$$F_1 = Y_1 \quad (4)$$

$\alpha$  es la constante de suavizamiento, que es estimada usando el criterio de menor Error Cuadrático Medio.

#### **Suavizamiento exponencial simple, excluyendo las últimas 5 observaciones**

Considerando los resultados obtenidos desde el Software R, el valor de nuestra constante de suavizamiento de nivel al excluir las 5 últimas observaciones es  $\alpha = 0,7144135$ , por ende, el modelo ajustado a nuestros datos es:

$$F_t = 0,7144135 * Y_t + (1 - 0,7144135) * F_t$$

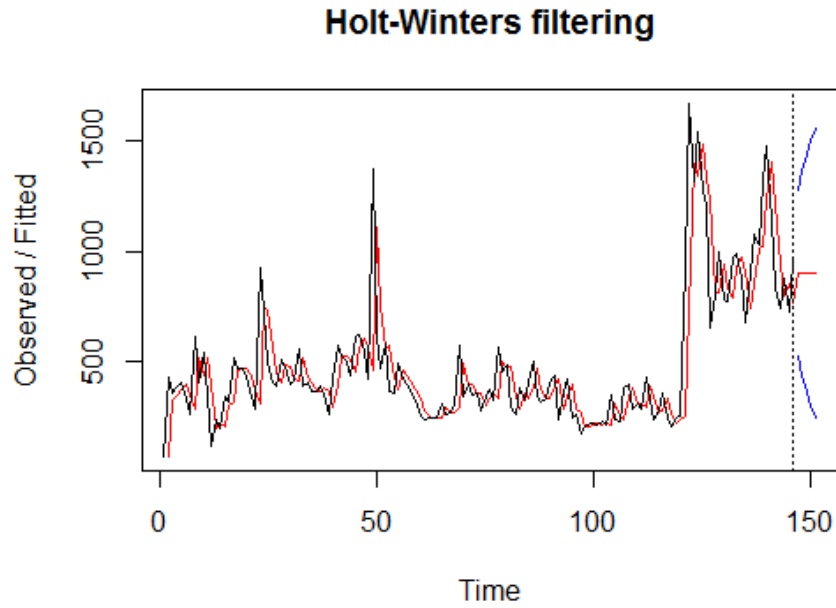


Figura 12: Suavizamiento Exponencial Simple, excluyendo las últimas 5 observaciones

Observación	Predicción
995	1277
919	1364
1242	1436
1095	1500
2293	1558

Cuadro 3: Predicción sobre las últimas 5 observaciones

Con *MAPE* de 0,2399771 sobre la muestra sin las últimas 5 observaciones (146 datos).

Con *MAPE* de 0,32284851 sobre las últimas 5 observaciones y sus respectivas predicciones.

### Suavizamiento exponencial simple, excluyendo las últimas 10 observaciones

Considerando los resultados obtenidos desde el Software R, el valor de nuestra constante de suavizamiento de nivel al excluir las 10 últimas observaciones es  $\alpha = 0,6886935$ , por ende, el modelo ajustado a nuestros datos es:

$$F_t = 0,6886935 * Y_t + (1 - 0,6886935) * F_t$$

Observación	Predicción
835	1594
738	1675
877	1743
725	1803
956	1857
995	1907
919	1954
1242	1998
1095	2040
2293	2079

Cuadro 4: Predicción sobre las últimas 10 observaciones

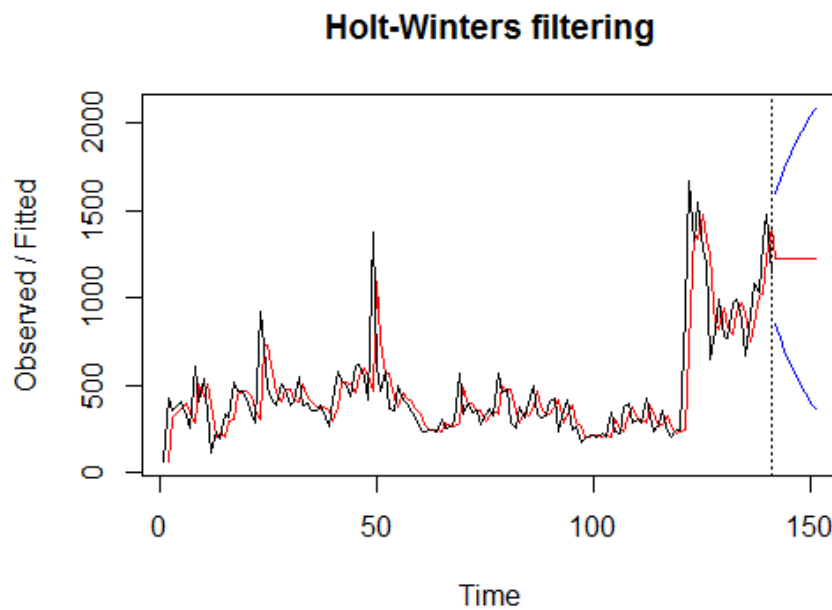


Figura 13: Suavizamiento Exponencial Simple, excluyendo las últimas 10 observaciones.

Con  $MAPE$  de 0,23977957 sobre la muestra sin las últimas 10 observaciones (141 observaciones). Con  $MAPE$  de 0,9203 sobre las últimas 10 observaciones.

### 6.1.2. Suavizamiento exponencial doble

El método de suavizamiento se utiliza en caso de que la serie presente nivel y tendencia, pero no estacionalidad. Está definido de la siguiente manera:

$$L_t = \alpha * Y_t + (1 - \alpha) * (L_{t-1} + b_{t-1}) \quad (5)$$

$$b_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1} \quad (6)$$

$$F_{t+m} = L_t + \beta_t * m \quad (7)$$

Donde  $L_t$  es la ecuación de suavizamiento de nivel,  $b_t$  es la ecuación de suavizamiento de la tendencia y  $F_{t+m}$  es la predicción al instante  $t + m$ .

Se pueden considerar las siguientes condiciones de inicialización:

$$L_1 = Y_1 \quad (8)$$

$$b_1 = Y_2 - Y_1 \quad (9)$$

### **Suavizamiento exponencial doble, excluyendo las últimas 5 observaciones**

Considerando los resultados obtenidos desde el Software R, luego de excluir las últimas 5 observaciones, el valor de nuestra constante de suavizamiento de nivel es  $\alpha = 0,8532131$  y el valor de la constante de suavizamiento de tendencia es  $\beta = 0,1401681$ , por ende, las ecuaciones de suavizamiento de nivel y tendencia para nuestros datos son:

$$L_t = 0,8532131 * Y_t + (1 - 0,8532131)(L_{t-1} + b_{t-1})$$

$$b_t = 0,1401681(L_t - L_{t-1}) + (1 - 0,1401681)b_{t-1}$$

Observación	Predicción
995	1336
919	1499
1242	1654
1095	1809
2293	1965

Cuadro 5: Predicción sobre las últimas 5 observaciones

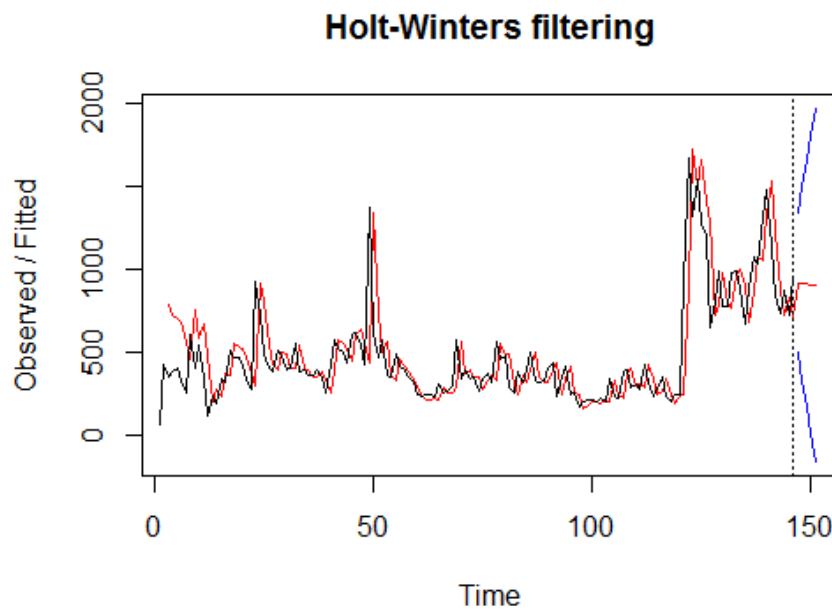


Figura 14: Suavizamiento Exponencial Doble, excluyendo las últimas 5 observaciones.

Con  $MAPE$  de 0,12795175 sobre la muestra sin las últimas 5 observaciones (146 datos). Con  $MAPE$  de 0,42013124 sobre las 5 últimas observaciones y sus respectivas predicciones.

### Suavizamiento exponencial doble, excluyendo las últimas 10 observaciones

Considerando los resultados obtenidos desde el Software R, luego de excluir las últimas 10 observaciones, el valor de nuestra constante de suavizamiento de nivel es  $\alpha = 0,82412251$  y el valor de la constante de suavizamiento de tendencia es  $\beta = 0,1511257$ , por ende, las ecuaciones de suavizamiento de nivel y tendencia para nuestros datos son:



Observación	Predicción
835	1650
738	1834
877	2013
725	2193
956	2377
995	2565
919	2757
1242	2955
1095	3157
2293	3365

Cuadro 6: Predicción sobre las últimas 10 observaciones

$$L_t = 0,82412251 * Y_t + (1 - 0,82412251)(L_{t-1} + b_{t-1})$$

$$b_t = 0,1511257(L_t - L_{t-1}) + (1 - 0,1511257)b_{t-1}$$

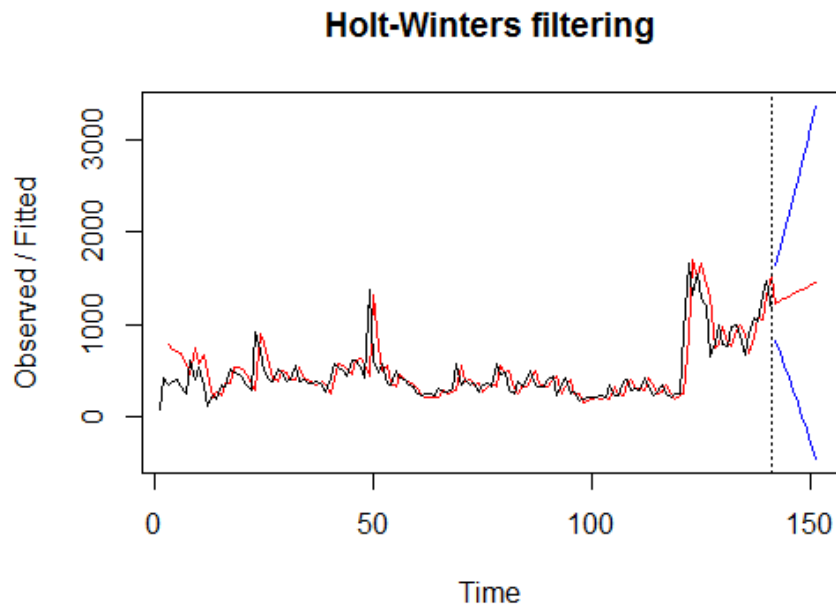


Figura 15: Suavizamiento Exponencial Doble, excluyendo las últimas 10 observaciones.

Con  $MAPE$  de 0,1347816 sobre la muestra sin las últimas 10 observaciones (141 observaciones).

Con *MAPE* de 1,4575 sobre las 10 últimas observaciones y sus respectivas predicciones.

## 6.2. Box-Jenkins

Este método tiene como propósito encontrar un modelo matemático que represente el comportamiento de nuestra serie de tiempo, con el fin de realizar predicciones.

### Modelos ARIMA

Los modelos *ARIMA* buscan explicar el comportamiento de una serie de tiempo a partir de observaciones pasadas de la misma serie y a partir de los errores pasados de predicción, siendo la notación de este modelo la siguiente:

$$ARIMA(p, d, q)$$

Donde:

- $p$ : Número de parámetros autorregresivos.
- $d$ : Número de diferenciaciones para convertir la serie a estacionaria.
- $q$ : Número de parámetros de medias móviles.

El modelo *ARIMA*( $p, d, q$ ) está representado por medio de la siguiente ecuación:

$$\phi(L) \nabla^d Y_t = \omega(L) \epsilon_t \quad (10)$$

Donde:

$$\nabla^d = (1 - L)^d \quad (11)$$

El modelo *ARIMA* se compone de dos partes, la parte autorregresiva (*AR*( $P$ )) que en este caso está representada por  $\phi(L)$  y la parte de medias móviles (*MA*( $q$ )) que está representada por  $\theta(L)$  y la parte de diferenciación está dada por  $\nabla^d$ .

### 6.2.1. Box-Jenkins, excluyendo últimas 5 observaciones

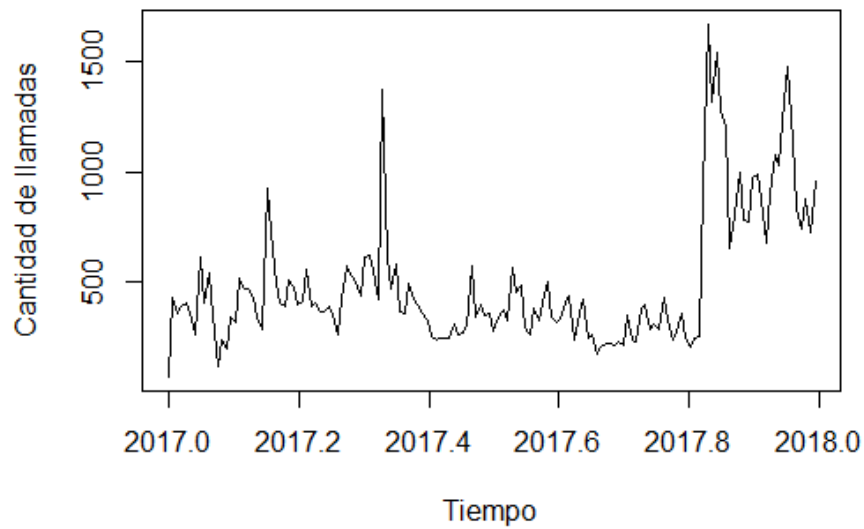


Figura 16: Serie original, excluyendo las últimas 5 observaciones

Considerando que nuestra serie original tiene varianza no constante, procederemos a aplicar logaritmo para graficar la serie resultante, además cabe mencionar que no se aprecia componente estacional, por ende, no es necesario aplicar el método *SARIMA* a nuestros datos.

Procederemos a analizar la Figura 17, para analizar el comportamiento de nuestros datos.

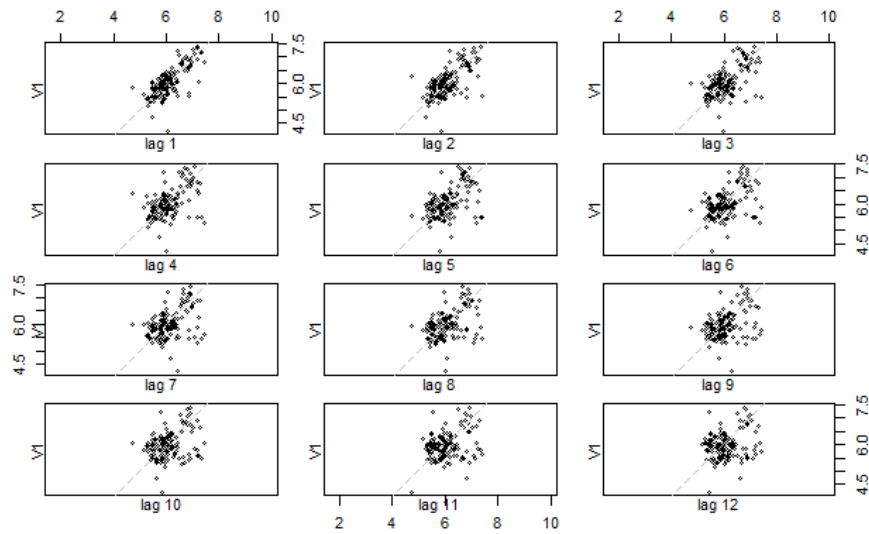


Figura 17: Gráfico lag

En la Figura 17 se observa gráficamente los Lag, siendo el primer elemento el gráfico de correlación entre la cantidad de llamadas del día de ayer vs. la cantidad de llamadas de antes de ayer, en la misma fila, pero la columna dos, se muestra el gráfico de la correlación entre la cantidad de llamadas del día de antes de ayer y del día anterior a antes de ayer y así sucesivamente. En general se aprecia correlación entre las cantidades de llamadas de cada día, que es precisamente lo que se busca, ya que , cumpliéndose esto, decimos que “la cantidad de llamadas de el día de hoy influye en la cantidad de llamadas que recibiremos mañana”.

Procederemos a analizar la FAC de los datos de manera gráfica, para corroborar si efectivamente la serie es estacionaria o no.

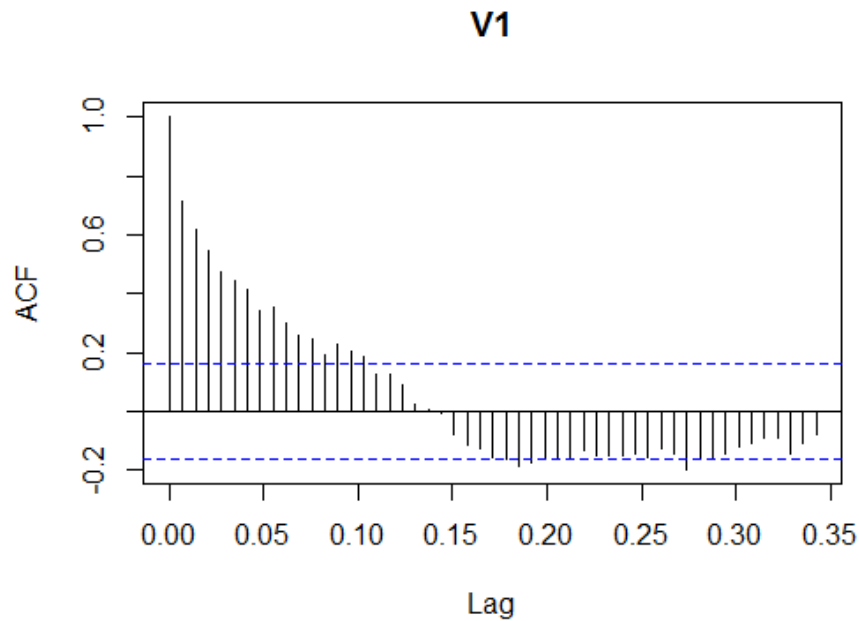


Figura 18: ACF

### Test de Dickey-Fuller

Considerando que existe cambio de nivel, decimos que nuestra serie no es estacionaria, para cercionarnos de esto, realizamos el test de Dickey-Fuller, considerando la siguiente hipótesis:

$H_0$ : la serie no es estacionaria (posee raíz unitaria)

$H_1$ : la serie es estacionaria (no posee raíz unitaria)

Luego de aplicar el test mediante el programa R, obtenemos un  $p - valor$  de 0,3463, por lo que no debemos rechazar la hipótesis nula y decimos que nuestra serie no es estacionaria.

Le aplicaremos logaritmo a la serie original y analizaremos su estacionariedad.

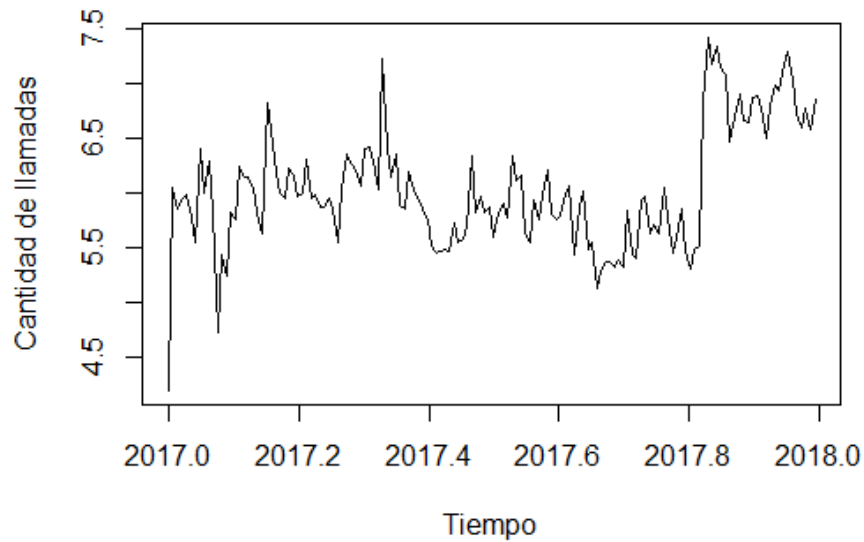


Figura 19: Serie con logaritmo aplicado

A pesar de aplicar logaritmo a los datos, la serie no es estacionaria, para corroborar si la serie es estacionaria o no, aplicaremos el test de Dickey Fuller, el cual nos entrega un  $p - valor$  de 0,4457, por ende no rechazamos la hipótesis nula y decimos que la serie con logaritmo no es estacionaria.

Como la serie original y la serie con logaritmo no son estacionarias, debemos diferenciar la serie con logaritmo, obteniendo los siguientes resultados:

Parámetro	Valor	p-valor
ar1	0.4932	0.0001
ar2	0.0580	0.5910
ma1	-0.8546	0.0000

Cuadro 7: Significancia de parámetros, excluyendo las últimas 5 observaciones

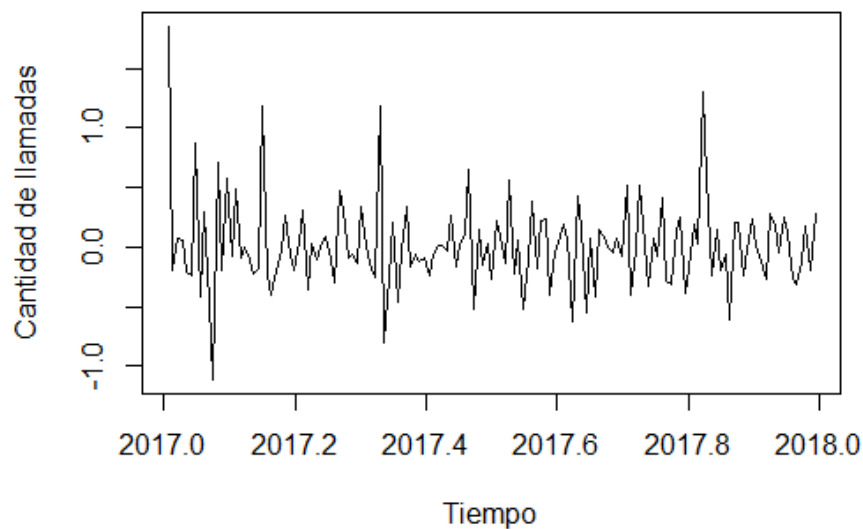


Figura 20: Serie con primera diferenciación

Al aplicar el test de Dickey-Fuller, nos entrega un  $p - valor$  de 0,01, por lo que rechazamos la hipótesis nula y decimos que la serie es estacionaria.

Luego usamos el comando `auto.arima` el cual nos sugiere un modelo  $ARIMA(2, 0, 1)$ , modelo que entrega un  $MAPE$  de 0,04124.

### Test de significancia individual

Consideramos el modelo entregado por el programa R y corroboraremos si efectivamente los parámetros entregados por el modelo son significativos.

$H_0$ : el parámetro no es significativo.

$H_1$ : el parámetro es significativo.

Compararemos algunos modelos:

Parámetro	Valor	p-valor
ar1	0.4852	0.00006
ma1	-0.8256	0.00000

Cuadro 9: Modelo, excluyendo las últimas 5 observaciones

Modelo	MAPE	AIC	BIC
0,1,1	4.042503	120.44	126.39
0,1,2	3.983220	119.40	128.33
1,1,0	4.082296	124.19	130.14
1,1,1	3.993664	117.46	126.39
1,1,2	3.986110	119.15	131.05

Cuadro 8: Comparación modelos ARIMA, excluyendo 5 observaciones

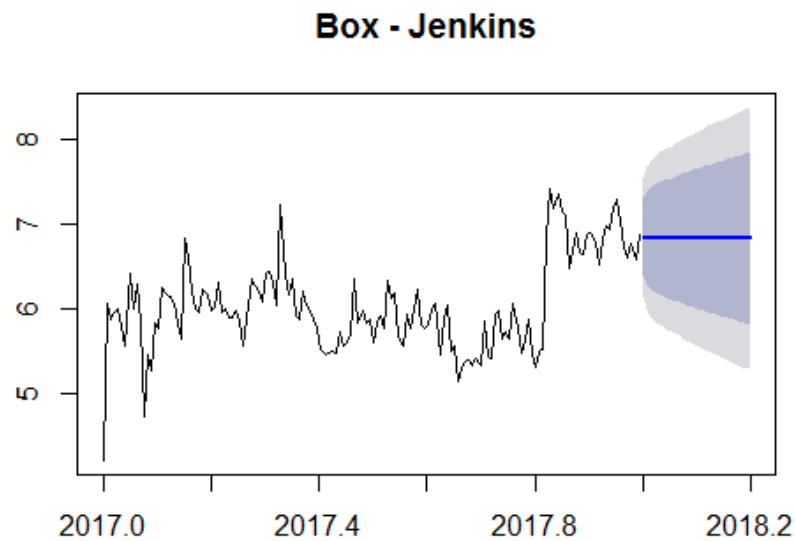


Figura 21: Serie ajustada

El gráfico del modelo ajustado, muestra la predicción en la línea azul, el intervalo de confianza del 95 % en celeste e intervalo de 80 % de confianza en gris. Además el modelo ajustado, se representa a través de la siguiente ecuación:

$$w_t = 0,4853w_{t-1} - 0,8256\varepsilon_{t-1} + \varepsilon_t$$

Ahora procederemos a analizar los residuos y su independencia:



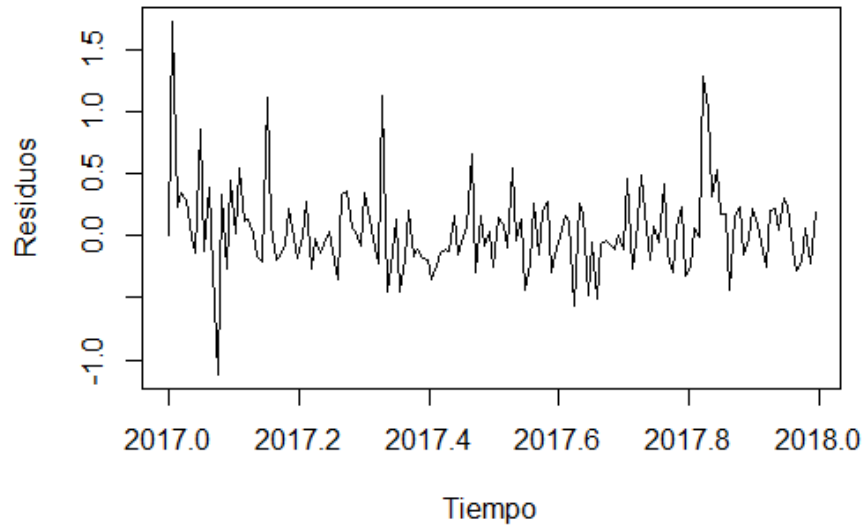


Figura 22: Residuos

El gráfico en cada punto muestra la diferencia entre la observación y el ajuste del modelo.

### Test Jarque Bera

Aplicamos el test Jarque Bera para analizar la normalidad de los residuos.

$H_0$ : los datos se distribuyen normal.

$H_1$ : los datos no se distribuyen normal.

El test nos entrega un  $p - valor < 2,2e - 16$ , por lo cual rechazamos la hipótesis nula y decimos que los datos no se distribuyen normal.

### Test de Ljung-Box

Aplicamos el test de Ljung-Box a los residuos para analizar su independencia, test que está sujeto a la siguiente hipótesis:

$H_0$ : los residuos son independientes

$H_1$ : los residuos no son independientes

El test nos entrega un  $p - valor$  de 0,8026, por ende no rechazamos la hipótesis nula y decimos que los residuos son independientes, a diferencia de los datos que son dependientes, que es precisamente el escenario ideal.

Ahora procederemos a realizar la comparación de las predicciones versus las últimas cinco observaciones que sacamos de la serie original:

Predicción con logaritmo	Predicción app.(aplicado exp)	Observación
6.842314	936	995
6.832393	927	919
6.827578	922	1242
6.825242	920	1095
6.824108	919	2293

Cuadro 10: Predicción sobre las últimas 5 observaciones

Con  $MAPE$  de 0,2169 sobre las últimas 5 observaciones y sus respectivas predicciones.

### 6.2.2. Box-Jenkins, excluyendo últimas 10 observaciones

Al igual que en el caso anterior, a pesar de quitar las últimas 10 observaciones, la serie sigue siendo no estacionaria, al aplicar logaritmo a esta serie y luego aplicar el test de Dickey-Fuller, la serie sigue siendo no estacionaria, por ende, pasamos directamente a la diferenciación de la serie con logaritmo, obteniendo el siguiente gráfico:

Parámetro	Valor	p-valor
ar1	0.4802	0.0004
ar2	0.0578	0.5999
ma1	-0.8428	0.0000

Cuadro 11: Significancia de parámetros, excluyendo las últimas 10 observaciones

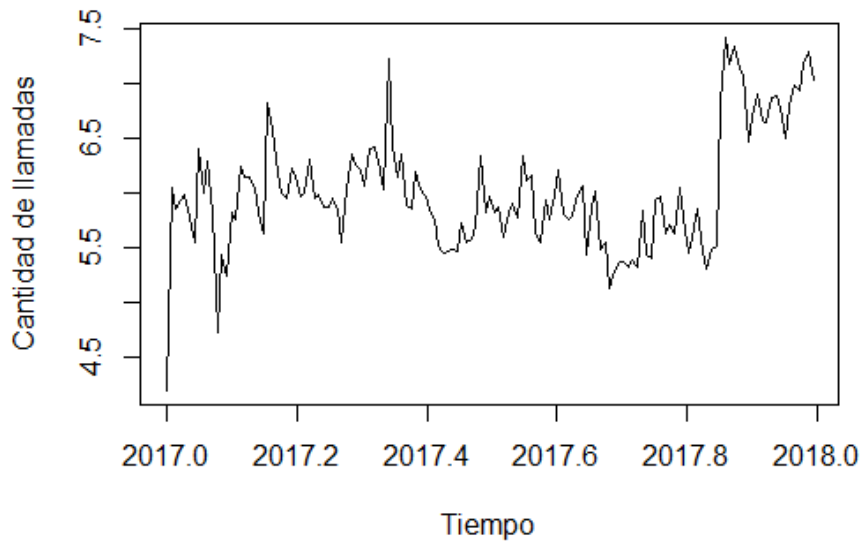


Figura 23: Serie con logaritmo aplicado

Luego de diferenciar una vez y después de aplicar el test de Dickey-Fuller, el cual entrega un  $p - valor$  de 0,01, rechazamos la hipótesis nula y decimos que la serie es estacionaria, por ende, a esta serie se le ajustará el modelo  $ARIMA(p, d, q)$ , para encontrar el modelo usaremos el comando `auto.arima`, el cual nos sugiere un modelo  $ARIMA(2, 0, 1)$ .

Consideramos el modelo entregado por el programa R y corroboraremos si efectivamente los parámetros entregados por el modelo son significativos o no.

Modelo	MAPE	AIC	BIC
0,1,1	4.070939	119.15	125.04
0,1,2	4.014492	118.46	127.29
1,1,0	4.11221	122.94	128.82
1,1,1	4.031922	116.94	125.76
1,1,2	4.022603	118.63	130.39

Cuadro 12: Comparación de modelos, excluyendo últimas 10 observaciones

Parámetro	Valor	p-valor
ar1	0.4693	0.0018
ma1	-0.8118	0.0000

Cuadro 13: Modelo, excluyendo las últimas 10 observaciones

### Test de significancia individual

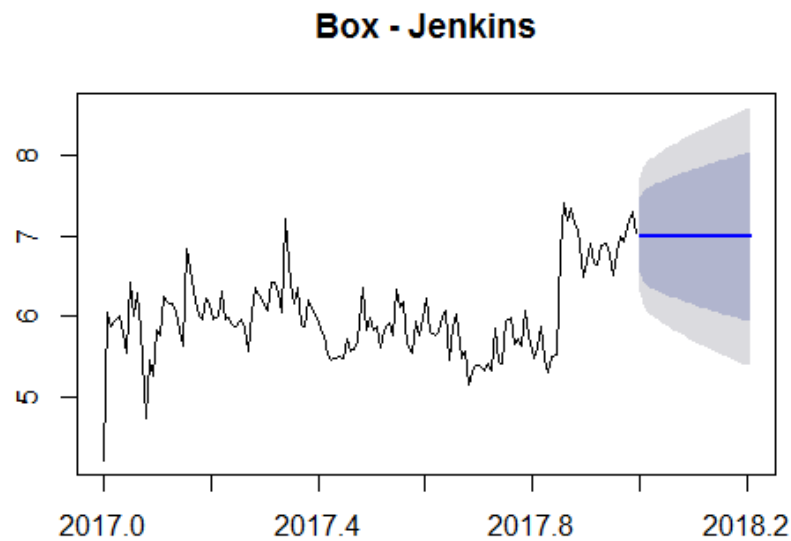


Figura 24: Serie ajustada, excluyendo las últimas 10 observaciones

El gráfico del modelo ajustado, muestra la predicción en la línea azul, el intervalo de confianza del 95 % en celeste e intervalo de 80 % de confianza en gris, además el modelo ajustado, se representa a través de la siguiente ecuación:

$$w_t = 0,4693w_{t-1} - 0,8119\epsilon_{t-1} + \epsilon_t$$

### Test Jarque Bera

Aplicamos el test Jarque Bera para analizar la normalidad de los residuos.

$H_0$ : los datos se distribuyen normal.

$H_1$ : los datos no se distribuyen normal.

El test nos entrega un  $p - valor < 2,2e - 16$ , por lo cual rechazamos la hipótesis nula y decimos que los datos no se distribuyen normal.

Procederemos a analizar y graficar los residuos del modelo y les aplicaremos test de independencia, el gráfico de los residuos del modelo es el siguiente:

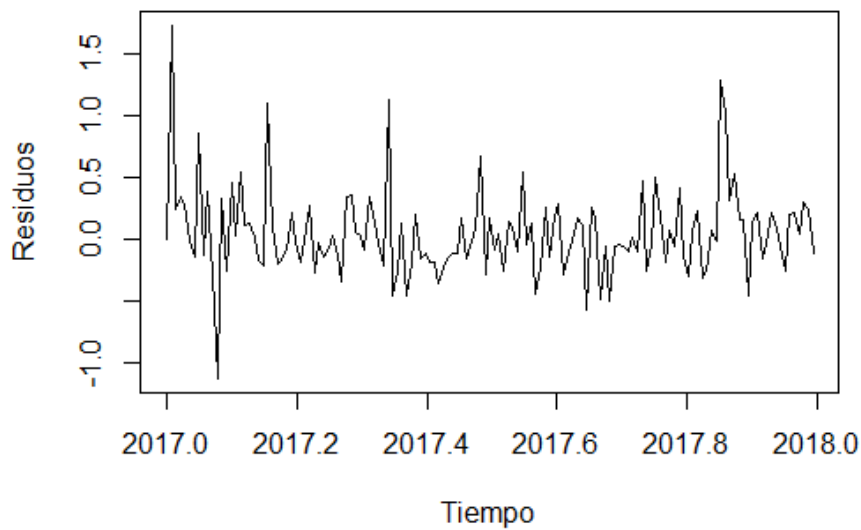


Figura 25: Residuos

### Test de Ljung-Box

Luego de aplicarle el test de Ljung-Box, nos entrega un  $p - valor$  de 0,4312, por ende no rechazamos hipótesis nula y decimos que los residuos de nuestro modelo son independientes, caso contrario a los datos que son dependientes.

Predicción con logaritmo	Predicción app.(aplicado exp)	Observación
6.842314	937	835
6.832393	927	738
6.827578	923	877
6.825242	921	725
6.824108	920	956
6.823558	919	995
6.823291	919	919
6.823161	919	1242
6.823099	918	1095
6.823068	918	2293

Cuadro 14: Predicción, excluyendo las últimas 10 observaciones

Con *MAPE* de 0,1836 sobre las últimas 10 observaciones y sus respectivas predicciones.

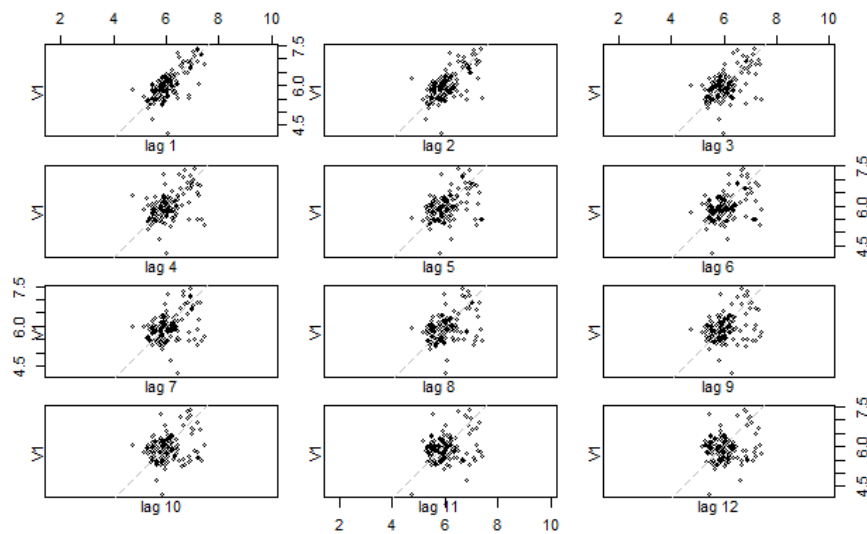


Figura 26: Gráfico lag

En la Figura 26 se observa gráficamente los Lag, siendo el primer elemento el gráfico de correlación entre la cantidad de llamadas del día de ayer vs. la cantidad de llamadas de antes de ayer, en la misma fila, pero la columna dos, se muestra el gráfico de la correlación entre la cantidad de llamadas del día de antes de ayer y del día anterior a antes de ayer y así sucesivamente. En general se aprecia correlación entre las cantidades de llamadas de cada día, que es precisamente lo que se busca, ya que , cumpliéndose esto, decimos que “la cantidad de llamadas de el día de hoy influye en la cantidad de llamadas que recibiremos mañana”.

### 6.3. Espacio Estado

#### Modelo de Nivel Local

$$y_t = \mu_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (12)$$

$$\mu_{t+1} = \mu_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2) \quad (13)$$

#### Modelo de Tendencia Local

$$y_t = \mu_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (14)$$

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2) \quad (15)$$

$$\beta_{t+1} = \beta_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (16)$$

#### Modelo Estructural Básico

$$y_t = \mu_t + \gamma_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (17)$$

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2) \quad (18)$$

$$\beta_{t+1} = \beta_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (19)$$

$$\gamma_{t+1} = - \sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t, \omega_t \sim N(0, \sigma_\omega^2) \quad (20)$$

#### 6.3.1. Espacio Estado, excluyendo últimas 5 observaciones

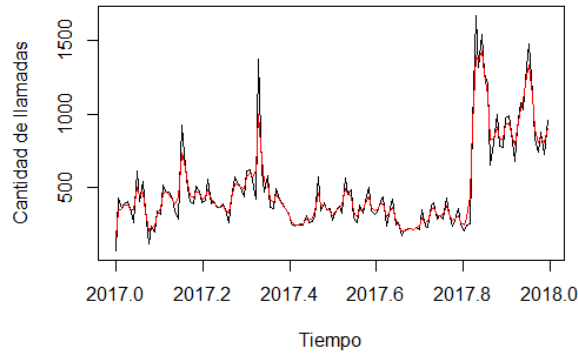


Figura 27: Muestra excluyendo últimas 5 observaciones, se observa el ajuste en rojo, predicción e intervalos de confianza del 90 %

Observación	Predicción
995	908
919	914
1242	919
1095	924
2293	929

Cuadro 15: Predicción sobre los últimas 5 observaciones

Con  $MAPE$  de 0,22079 para las últimas 5 observaciones y sus respectivos ajustes.

### 6.3.2. Espacio Estado, excluyendo últimas 10 observaciones

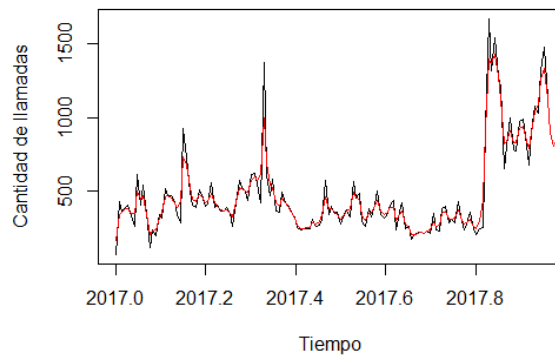


Figura 28: Muestra excluyendo últimas 10 observaciones, se observa el ajuste en rojo, predicción e intervalos de confianza del 90 %

Observación	Predicción
835	1231
738	1239
877	1246
725	1254
956	1262
995	1269
919	1277
1242	1284
1095	1292
2293	1299

Cuadro 16: Predicción sobre las últimas 10 observaciones



Con  $MAPE$  de 0,3935 para las últimas 10 observaciones y sus respectivos ajustes.

## 7. Comparación

Ahora procederemos a comparar los MAPE anteriormente calculados, en el Cuadro 17 se comparan los MAPE sobre las observaciones que dejamos fuera de la serie original y en el Cuadro 18 se comparan los MAPE sobre la serie sin las últimas observaciones.

Método	Sobre últimas 5 obs.	Sobre últimas 10 obs.
Suavizamiento exp. Simple	0.239977098	0.920329608
Suavizamiento exp. Doble	0.420131244	1.457542832
Box - Jenkins	0.216936581	0.183644779
Espacio Estado	0.220792116	0.393575379

Cuadro 17: Comparación MAPE, sobre últimas 5 y 10 observaciones

En el Cuadro 17, observamos que el método que menor error entrega sobre las últimas 5 y 10 observaciones es el método de Box-Jenkins.

Método	Sin últimas 5 obs.	Sin últimas 10 obs.
Suavizamiento exp. Simple	0.2399771	0.23977957
Suavizamiento exp. Doble	0.12795175	0.1347816
Box - Jenkins	3.993664	3.993664
Espacio Estado	0.091427578	0.12832688

Cuadro 18: Comparación MAPE, excluyendo últimas 5 y 10 observaciones

Analizando el Cuadro 18, tenemos que el método que mejor se ajusta a la serie sin las últimas 5 y 10 observaciones es la metodología de Espacio Estado.

## 8. Conclusión

Considerando la tabla de comparación de los métodos de predicción, podemos decir en lo que concierne a la serie que excluye las últimas 5 observaciones, el modelo con mejor desempeño es del método Espacio Estado, en el caso de las 5 observaciones excluidas y sus respectivos ajustes, al igual que en la serie donde se excluyen las últimas 10 observaciones.

Por otro lado, la metodología que presenta mejor predicción es el método de Box-Jenkins, seguido por el método de Espacio Estado. En ambos casos (con la serie que excluye las últimas 5 observaciones y la serie que excluye las últimas 10 observaciones).

Por tanto en lo que respecta al ajuste, el mejor método es Espacio Estado. Por otro lado, considerando los resultados obtenidos, la mejor predicción es entregada por la metodología de Box-Jenkins, que es lo que buscamos en nuestro estudio. Por lo demás, cabe mencionar que los métodos con peor desempeño o menos efectivos son los métodos de Holt-Winters.

## 9. Códigos en R

---

### SUAVIZAMIENTO EXPONENCIAL

---

---

Suavizamiento exponencial simple con 5 observaciones menos

---

```
serie<-read.table("normalesmenoscinco.txt",header=F)
normales<-ts(serie,frequency=146,start=c(2017,1))
plot(normales)
(m<-HoltWinters(serie,beta=F,gamma=F))
p<-predict(m,5,prediction.interval=T)
plot(m,p)
```

---

Suavizamiento exponencial simple con 10 observaciones menos

---

```
serie1<-read.table("normalesmenosdiez.txt",header=F)
normales<-ts(serie1,frequency=141,start=c(2017,1))
plot(normales)
(m<-HoltWinters(serie1,beta=F,gamma=F))
p<-predict(m,10,prediction.interval=T)
plot(m,p)
```

---

Suavizamiento exponencial doble con 5 observaciones menos

---

```
serie2<-read.table("normalesmenoscinco.txt",header=F)
normales<-ts(serie2,frequency=146,start=c(2017,1))
plot(normales)
(m<-HoltWinters(serie2,gamma=F))
p<-predict(m,5,prediction.interval=T)
plot(m,p)
```

---

Suavizamiento exponencial doble con 10 observaciones menos

---

```
serie3<-read.table("normalesmenosdiez.txt",header=F)
normales<-ts(serie3,frequency=141,start=c(2017,1))
```

```
plot(normales)
(m<-HoltWinters(serie3,gamma=F))
p<-predict(m,10,prediction.interval=T)
plot(m,p)
```

---

BOX-JENKINS

---



---

Box-Jenkins, excluyendo últimas 5 observaciones

---

```
serie<-read.table("normalesmenoscinco.txt",header=F)
pred<-ts(serie, start = c(2017,1),frequency = 146)
plot(pred, xlab="Tiempo", ylab="Cantidad de llamadas")
predlog<-log(pred)
plot(predlog, xlab="Tiempo", ylab="Cantidad de llamadas")
adf.test(predlog)
```

---

Test de Dickey-Fuller

---

```
adf.test(pred)
nsdiffs(predlog)
```

---

Primera diferenciación

---

```
c<-diff(predlog)
c
plot(c, xlab="Tiempo", ylab="Cantidad de llamadas")
adf.test(c)
```

---

Ajustar modelo autoarima

---

```
a=auto.arima(c)
```

```

summary(a)
fit <- Arima(predlog,order=c(1,1,2))
summary(fit)
plot(fit)
tsdisplay(residuals(fit))
res<-residuals(fit)
plot(res, xlab="Tiempo", ylab="Residuos")
Box.test (res,lag = 1, type = "Ljung")
plot(forecast(fit,h=30), main='Box - Jenkins')

acf(res)
pacf(res)

#Test Jarque Bera
jarque.bera.test(res)

#Cargar la librería fpp
library(fpp)
library(forecast)
#Realizaremos el gráfico lag, para analizar visualmente
#si la serie es estacionaria
lag.plot(predlog,lags=1,do.lines=F)

#veamos la fac y facp de los datos
acf(predlog,50)

-----
BOX-JENKINS
-----

-----
Box-Jenkins, excluyendo últimas 10 observaciones
-----

serie<-read.table("normalesmenosdiez.txt",header=F)
pred<-ts(serie, start = c(2017,1),frequency = 141)
plot(pred, xlab="Tiempo", ylab="Cantidad de llamadas")
predlog<-log(pred)
plot(predlog, xlab="Tiempo", ylab="Cantidad de llamadas")
adf.test(predlog)

```

---

Test de Dickey-Fuller

---

```
adf.test(pred)
nsdiffs(predlog)
```

---

Primera diferenciación

---

```
c<-diff(predlog)

plot(c, xlab="Tiempo", ylab="Cantidad de llamadas")
adf.test(c)
```

---

Ajustar modelo autoarima

---

```
a=auto.arima(c)
summary(a)
fit <- Arima(predlog,order=c(1,1,2))
summary(fit)
plot(fit)

tsdisplay(residuals(fit))
res<-residuals(fit)
plot(res, xlab="Tiempo", ylab="Residuos")
Box.test (res,lag = 1, type = "Ljung")
plot(forecast(fit,h=30), main='Box - Jenkins')

acf(res)
pacf(res)

#Test Jarque Bera
jarque.bera.test(res)

#cargar la librería fpp
library(fpp)
library(forecast)

lag.plot(predlog,lags=1,do.lines=F)
```

```
#Veamos la fac y facp de los datos
acf(predlog,50)
```

```
-----
ESPACIO ESTADO
-----
```

```
-----
Espacio Estado, excluyendo últimas 5 observaciones
-----
```

```
library(fpp)
library(forecast)
library(tseries)
library(KFAS)

serie<-read.table("normalesmenoscinco.txt",header=F)
normales<-ts(serie,start=c(2017,1),frequency=146)
plot(normales, xlab="Tiempo", ylab="Cantidad de llamadas")

model_normales<- SSMModel(normales ~ SSMtrend(2,
Q = list(matrix(NA),matrix(0))), H = matrix(NA))
model_normales

fit_structural <- fitSSM(model_normales,
inits = c(0, 0,0), method = "BFGS")
fit_structural$model["Q"]
fit_structural$model["H"]

out_gaussian <- KFS(fit_structural$model)
out_gaussian

summary(out_gaussian)

lines(out_gaussian$alphahat[,1],col=2)

model_normales1<- fitSSM(model_normales,
inits = c(0, 0,0), method = "BFGS")$model
model_normales1
p<-predict(model_normales1,n.ahead=5,
interval='prediction',level=0.9)
p
```



```
-----  
Espacio Estado, excluyendo últimas 10 observaciones  
-----
```

```
serie<-read.table("normalesmenosdiez.txt",header=F)  
normales<-ts(serie,start=c(2017,1),frequency=141)  
plot(normales, xlab="Tiempo", ylab="Cantidad de llamadas")
```

```
model_normales<- SSMModel(normales ~ SSMtrend(2,  
Q = list(matrix(NA),matrix(0))), H = matrix(NA))  
model_normales
```

```
fit_structural <- fitSSM(model_normales,  
inits = c(0, 0,0), method = "BFGS")  
fit_structural$model["Q"]  
fit_structural$model["H"]
```

```
out_gaussian <- KFS(fit_structural$model)  
out_gaussian
```

```
summary(out_gaussian)
```

```
lines(out_gaussian$alphahat[,1],col=2)
```

```
model_normales1<- fitSSM(model_normales,  
inits = c(0, 0,0), method = "BFGS")$model  
model_normales1  
p<-predict(model_normales1,n.ahead=10,  
interval='prediction',level=0.9)  
p
```

## Referencias

- [1] MAKRIDAKIS S., WHEELWRIGHT S.C. & HYNDMAN R.J., «Forecasting methods and applications. John Wiley & Sons», 2008.
- [2] EZEQUIEL URIEL & AMADO PEIRÓ, «Introducción al análisis de series temporales», 2000.
- [3] GONZALO RÍOS & CARLOS HURTADO, «Series de tiempo», Tópicos en minería de datos, Departamento de Ciencias de la Computación, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, 2008.
- [4] J. DURBIN & S.J.KOOPMAN, «Time Series Analysis by State Space Methods», 2012.
- [5] GEORGE E.P.BOX, GWILYM M.JENKINS, GREGORY C.REINSEL & GRETA M.LJUNG, «Time Series Analysis Forecasting and Control».
- [6] WALTER ENDERS, «Applied econometric time series», 2002.
- [7] JIMÉNEZ, J., «Pronóstico de demanda de llamadas en los Call Center, utilizando redes neuronales artificiales». Tesis de pregrado en Ingeniería Industrial y de Sistemas. Universidad de Piura. Facultad de Ingeniería. Programa Académica de Ingeniería Industrial y de Sistemas. Piura, Perú, 2013.