

Análisis de Series Temporales: Tarea 3

Sergio García Prado

sergio.garcia.prado@alumnos.uva.es

3 de enero de 2019

- **Archivo:** `weight-loss.csv`

- **Serie:** Frecuencia de búsquedas para la palabra clave “Weight loss” a través del buscador *Google* por meses, desde *Enero de 2004* hasta *Diciembre de 2018*. Los valores han sido estandarizados en el rango [0, 100].

1. Etapa de identificación

1.1. Contexto

Tal y como se indica al comienzo de este documento, en este trabajo se va a trabajar con la serie temporal referida a la frecuencia de búsqueda de la palabra clave “Weight loss” (a nivel mundial) a través del buscador *Google*. Se ha escogido esta serie para el trabajo por su estructura estacional claramente marcada. Se cree que dicha estructura tiene un alto grado de relación con un índice sobre la preocupación de la población por su peso a lo largo del tiempo.

Para evitar problemas de privacidad, los datos se proporcionan estandarizados en el rango [0, 100], lo cual elimina la escala de los mismos y únicamente permite estudiar la estructura estocástica de la serie. Esto no es un problema para el análisis que se realizará en este trabajo, dado que precisamente el objetivo del mismo es el de analizar la estructura de una serie temporal, siguiendo la metodología de *Box-Jenkins*.

En cuanto al particionamiento de los datos, estos se proporcionan en agrupaciones mensuales. Dado que se tiene información desde *Enero de 2004* hasta *Diciembre de 2018*, es decir, un total de *15 años*, lo cual suma $15 * 12 = 180$ observaciones en total. Con esta cantidad de observaciones, se cree que se podrá construir un modelo *SARIMA* (*ARIMA* con estacionalidad) de manera adecuada.

Una vez introducido el contexto de los datos pertenecientes a la serie que se analizará, lo siguiente es empezar a describir la misma a nivel de su estructura estocástica. Tras describir la misma, se procederá a realizar las diferenciaciones pertinentes hasta conseguir que esta sea estacionaria. Una vez se haya conseguido transformar la serie en estacionaria, se tratarán de identificar los parámetros de la parte autoregresiva y de la parte de media móvil, tanto de la dependencia entre observaciones a nivel serial (cada observación con las anteriores), como de la dependencia estacional (cada observación con las anteriores dentro de su periodo estacional). Tras dicha descripción, se propondrán un conjunto de modelos *SARIMA*. En la Sección 2 se prodecerá al ajuste de dichos modelos a los datos. Posteriormente, en la Sección 3 serán descartados aquellos modelos que no puedan validarse por su excesiva falta de ajuste, sobre ajuste, parámetros no significativos, etc. De entre los modelos válidos, se seleccionará aquel cuyo ajuste sea el más próximo a los datos, lo cual se comprobará mediante distintas técnicas. Finalmente, en la Sección 4 se realizará una predicción para el próximo año (2019) sobre los valores esperados por el modelo seleccionado.

La metodología que se ha expuesto en el párrafo anterior se corresponde con la propuesta por *Box-Jenkins* para series temporales basada en ajuste de modelos *ARIMA*. En el documento, se sigue un enfoque en paralelo en lugar de iterativo para la búsqueda del mejor modelo para facilitar la interpretación y la organización del mismo. Esta es la única modificación que se ha llevado a cabo respecto de la metodología original.

1.2. Análisis Descriptivo

Tras la descripción de la metodología, se va a comenzar con la descripción de la serie temporal. Para ello, nos vamos a apoyar en los gráficos de la Figura 1, a partir de los cuales se puede tener una perspectiva completa acerca de la serie. A través de ella se puede ver el gráfico de la serie, el correlograma, el correlograma parcial, el periodograma y el diagrama de dispersión *rango-media*.

En el gráfico de la serie de la Figura 1 se puede apreciar la evolución temporal de los valores a lo largo de los 15 años, separados en observaciones mensuales. En dicha representación destacan dos características de la serie temporal sobre el resto. Estas son: (1) la marcada estructura estacional de periodo 12 (anual) de la serie, que sigue la misma forma en todas las estacionalidades (años). Esto es un fuerte crecimiento durante el primer mes (Enero), posteriormente se produce un suave decrecimiento hasta el tercer cuarto del año, para producirse un fuerte decrecimiento en torno a los meses de Septiembre - Octubre. Esta estructura estacional es coherente con el fenómeno conocido como *Operación Bikini*, que consiste en la preocupación por estar en buena forma física durante los meses de verano. Esta preocupación comienza en torno a principios de año y se mantiene hasta los meses de verano. Puesto que una vez pasados dichos meses, la forma física deja de estar comprometida, la preocupación por la pérdida de peso de la población también disminuye. (2) el cambio de nivel que se produce entre el año 2008 y el año 2009. Durante este periodo se produce un cambio drástico en el nivel de la serie. Parece que durante el final del año 2008 no se produjo el fenómeno esperado de un fuerte decrecimiento que sí se produce durante el resto de años. Este cambio en el nivel de la serie pudo deberse a distintos factores. Uno de ellos pudo ser la crisis económica cuyas evidencias comonezaron en torno a dicho año. Si se confirmase que las razones del aumento del nivel en las búsquedas del término *Weight loss* fueron debidas a la crisis económica, una interpretación para la misma podría ser la siguiente: Con el aumento del riesgo en la estabilidad financiera, la población aumentó también su preocupación en su apariencia física, lo cual se ha mantenido hasta la actualidad. Otro de los factores de peso pudo haber sido el acercamiento de la tecnología y redes sociales a la mayoría de la población, que hasta entonces se había mantenido alejada de ella. Dado que muchas redes sociales permiten la posibilidad de incluir imágenes personales que otras personas pueden juzgar, el índice de preocupación por el nivel de forma física puede haberse incrementado en los últimos años debido a dicha razón. Sin embargo, estas interpretaciones queda fuera del objetivo de este trabajo.

En cuanto al correlograma de la serie que se muestra en la Figura 1, se puede apreciar la componente estacional de periodo 12 en la estructura de correlaciones. Destacan sobre el resto los retardos de la forma $i \bmod 12 = 0$, estos son los retardos 12, 24, 36, Estos se relacionan entre si presentando un decrecimiento lineal, por lo que son indicativo de que la serie no es estacionaria. Por lo tanto, tendremos que llevar a cabo al menos una diferenciación estacional para conseguir estacionarizar la serie. También llama la atención la gran cantidad de correlaciones con valores significativos, por lo que la tendencia podría estar ocultando algún otro comportamiento no visible a simple vista. Por lo tanto, la realización de una diferenciación regular también podría ser una buena estrategia para tratar de comprender en mayor medida la estructura de correlaciones de la serie. Como se verá en el siguiente párrafo, estas interpretaciones se ven reflejadas en el correlograma parcial.

En el correlograma parcial de la serie de la Figura 1 se representan las correlaciones entre observaciones de la serie, tratando de eliminar de estas la relación procedente de otros retardos. Es decir, el correlograma parcial trata de representar de manera aislada la correlación entre una observación y la del k -ésimo retardo posterior, eliminando la influencia del resto. En este caso destacan sobre el resto los retardos 1, 12 y 13. Se cree que el retardo 1 destaca sobre el resto debido a la tendencia de la serie mientras que el retardo 12 se debe a la estacionalidad de la misma. También se piensa que el retardo 13 es un reflejo del 1, en la estacionalidad anterior, de ahí la razón de que destaque de tal manera.

En cuanto al periodograma de la serie que se muestra en la Figura 1, se puede apreciar que el primer armónico recoge gran parte de la variabilidad de la serie, lo cual de nuevo vuelve a indicar que la tendencia de la misma puede estar ocultando información sobre la estructura estocástica de la serie. Por lo tanto, se cree que una diferenciación permitiría visualizar en mayor medida el comportamiento de la misma. También destacan (aunque de una manera mucho menos pronunciada) los armónicos de la forma $i + 1/12$ con $i \in 1, 2, \dots, 6$, esto es $1/12, 2/12, \dots, 6/12$, lo cual es otro argumento de peso en favor de la componente estacional de periodo 12 (anual) de la serie.

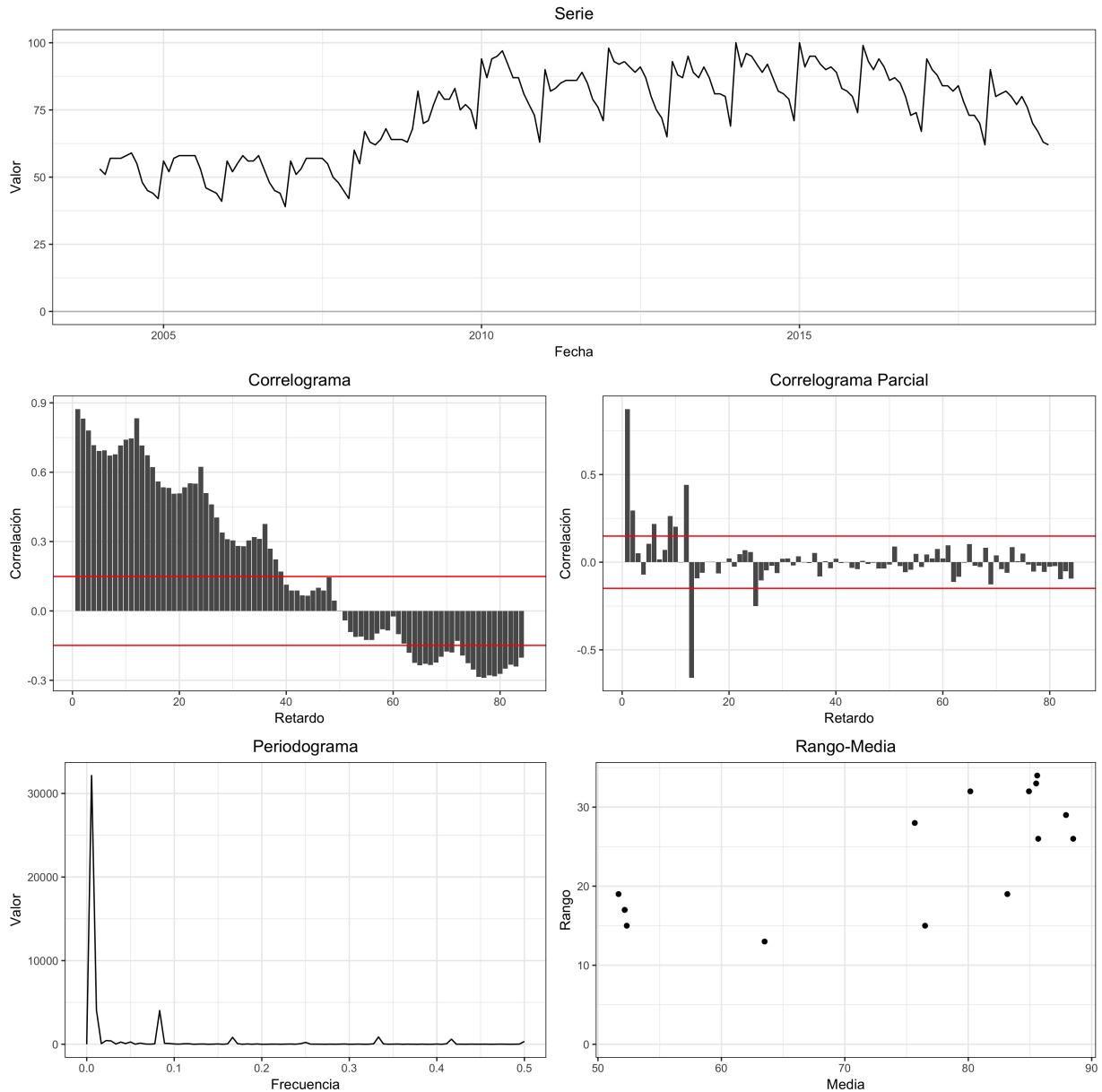


Figura 1: Gráfico de la serie, Correlograma, Correlograma Parcial, Periodograma y diagrama de dispersión *rango-media* para la serie **weightloss**.

En cuanto al diagrama de dispersión *rango-media* de la Figura 1, parece que existe una leve relación entre el nivel de la serie y la dispersión del mismo. Este fenómeno podría requerir de alguna transformación de estabilización de varianza, como las de la forma *Box-Cox*. La razón de ello es que los modelos que se ajustarán a la serie temporal requieren de estacionaridad en la serie (entre otros requisitos, la varianza debe ser constante a lo largo de la serie). Sin embargo, tal y como se verá posteriormente, las diferenciaciones eliminarán la relación entre nivel y dispersión, por lo que es algo de lo que no nos preocuparemos en gran medida.

1.3. Diferenciaciones

Anteriormente, se ha realizado un análisis descriptivo acerca de la estructura de la serie **weightloss**, tras el cual se llegó a la conclusión de que la serie no presentaba la propiedad de estacionaridad (nivel y estructura de correlaciones constante para todas las observaciones). Por tanto, se indicó que para llegar a una serie estacionaria, la solución sería encontrar una diferenciación o varias diferenciaciones que transformen la serie en estacionaria.

Tal y como se indicó anteriormente, para conseguir la propiedad de estacionaridad dichas diferenciaciones podrían ser de dos tipos: (1) diferenciación regular para reducir la componente de tendencia de la serie, y (2) diferenciación estacional de periodo 12 para reducir las correlaciones entre observaciones de un mismo índice estacionalidad. Entonces, la estrategia a seguir será la de probar entre distintas combinaciones de diferenciaciones (de orden reducido, ya que de no ser así entonces se producirían efectos tales como el aumento de la varianza de la serie).

Por tanto, para elegir el orden de diferenciación se han probado distintas alternativas. En la Tabla 1 se incluye una relación entre la varianza resultante de las observaciones de la serie y el orden de diferenciación de la misma. Como se puede apreciar, la serie original tiene una varianza de 258,67. Sin embargo, la de la serie tras aplicar una diferenciación regular y otra estacional es únicamente de 14,07.

Serie	Varianza
X_t	258.67
∇X_t	62.92
$\nabla_{12}X_t$	49.09
$\nabla_{12}^2X_t$	86.28
$\nabla\nabla_{12}X_t$	14.07
$\nabla\nabla_{12}^2X_t$	42.11
$\nabla^2\nabla_{12}X_t$	34
$\nabla^2\nabla_{12}^2X_t$	99.44

Tabla 1: Relación entre distintos órdenes de diferenciación y varianza de la serie resultante.

Tanto por la reducida varianza, como por la propiedad de estacionaridad, finalmente escogeremos los grados de diferenciación 1 para la componente regular y 1 para la componente estacional. Sin embargo, a continuación se van a estudiar un poco más en detalle los resultados de la serie tras la aplicación de distintas diferenciaciones: En la Subsección 1.3.1 se analiza la serie diferenciada regularmente y en la Subsección 1.3.2 se analiza la serie diferenciada estacionalmente. Por último, en la Subsección 1.3.3 se muestra la serie diferenciada tanto regular como estacionalmente. Tal y como se ha indicado anteriormente, esta será la serie con la que se continuará el análisis y posterior ajuste.

1.3.1. Diferenciación regular

La diferenciación regular consiste en la aplicación del operador ∇ una vez a la serie objetivo. Esto es equivalente a multiplicar la misma por el término $(1 - B)$, donde B consiste en el operador *backward*, que genera como resultado la observación anterior. En la Ecuación 1 se muestra la representación de la diferenciación regular. Esto consiste en substraer el valor de la observación actual a la anterior. Tras aplicar esta operación a todas las observaciones de la serie, se consigue la eliminación de una tendencia de carácter lineal. Nótese que para hacer esto, se están perdiendo 1 observación.

$$\begin{aligned}\nabla X_t &= (1 - B) \cdot X_t \\ &= X_t - X_{t-1}\end{aligned}\tag{1}$$

En la Figura 2 se muestran el gráfico de la serie, el correlograma, el correlograma Parcial, el periodograma y el diagrama de dispersión *rango-media* para la serie tras la realización de una diferenciación regular.

Tal y como se puede apreciar en el gráfico de la serie, se ha eliminado la tendencia de la misma. Ahora la media está en torno al valor 0, sobre el cual oscila. Sin embargo, se puede apreciar que la serie sigue teniendo una estructura estacional, que la diferenciación no ha podido eliminar. En el correlograma se puede apreciar que dicha estructura de estacionalidad (con las correlaciones múltiples de 12 claramente marcadas) que estas presentan un decrecimiento de carácter lineal, lo cual indica que la serie no es estacionaria. En el correlograma parcial se puede apreciar la reducción de la correlación del primer retardo, lo cual era algo obvio tras la diferenciación regular. En cuanto al periodograma, ahora se pueden apreciar de una manera

muy marcada los armónicos significativos, los cuales siguen la forma $i/12$. Este es otro de los argumentos en favor de la diferenciación estacional, que se realizará en la Subsubsección 1.3.2. En cuanto al gráfico de dispersión *rango-media*, se puede comprobar que ahora se ha conseguido controlar en mayor medida la relación entre el nivel y la dispersión, es decir, se ha estabilizado en mayor medida la varianza. Sin embargo, esto se conseguirá en mayor media en las diferenciaciones posteriores.

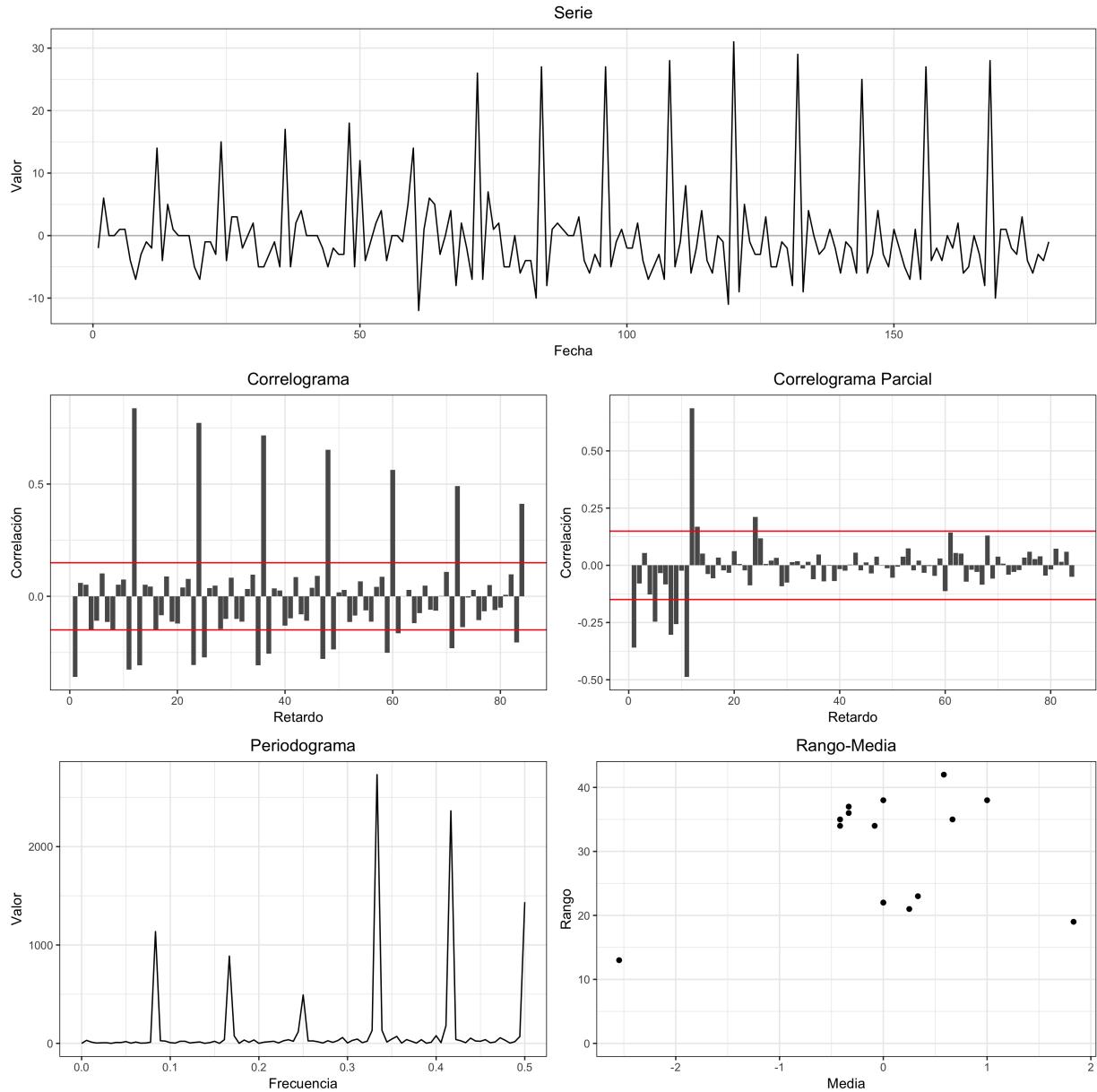


Figura 2: Gráfico de la serie, Correlograma, Correlograma Parcial, Periodograma y diagrama de dispersión *rango-media* para la serie **weightloss** tras la realización de una diferenciación regular.

La aplicación de la una diferenciación regular a la serie **weightloss** ha repercutido de manera positiva sobre la misma desde el punto de vista de la estacionaridad (ahora se encuentra más cerca de cumplir dicha propiedad). En la Tabla 1 se indica que el valor de la varianza es 62,92, lo cual es una drástica reducción en la dispersión de la serie. Sin embargo, todavía no puede ser considerada estacionaria. A continuación probaremos la alternativa estacional y estudiaremos los resultados.

1.3.2. Diferenciación estacional

La diferenciación estacional se basa en las mismas ideas que la diferenciación regular descrita anteriormente. Sin embargo, en este caso se utiliza una notación de subíndices para indicar el retardo en que aplicar

la diferenciación. Esto es, en lugar de substraer de la observación actual la inmediatamente anterior, se substraen la anterior respecto referida al mismo índice estacional. En nuestro caso, esto se refiere a substraer el valor del mismo mes del año anterior, esto es la observación 12 posiciones hacia atrás. En la Ecuación 2 se muestra la expresión de la diferenciación que se acaba de describir. Tal y como se puede apreciar, esta corresponde con la substracción de la observación 12 posiciones hacia atrás. Si se aplica esta operación a todas las observaciones de la serie, entonces se consigue la serie diferenciada estacionalmente. Nótese que para hacer esto, se están perdiendo 12 observaciones.

$$\begin{aligned}\nabla_{12}X_t &= (1 - B^{12}) \cdot X_t \\ &= X_t - X_{t-12}\end{aligned}\tag{2}$$

En la Figura 3 se muestran el gráfico de la serie, el correlograma, el correlograma parcial, el periodograma y el diagrama de dispersión *rango-media* para la serie tras la realización de una diferenciación estacional (12 retardos).

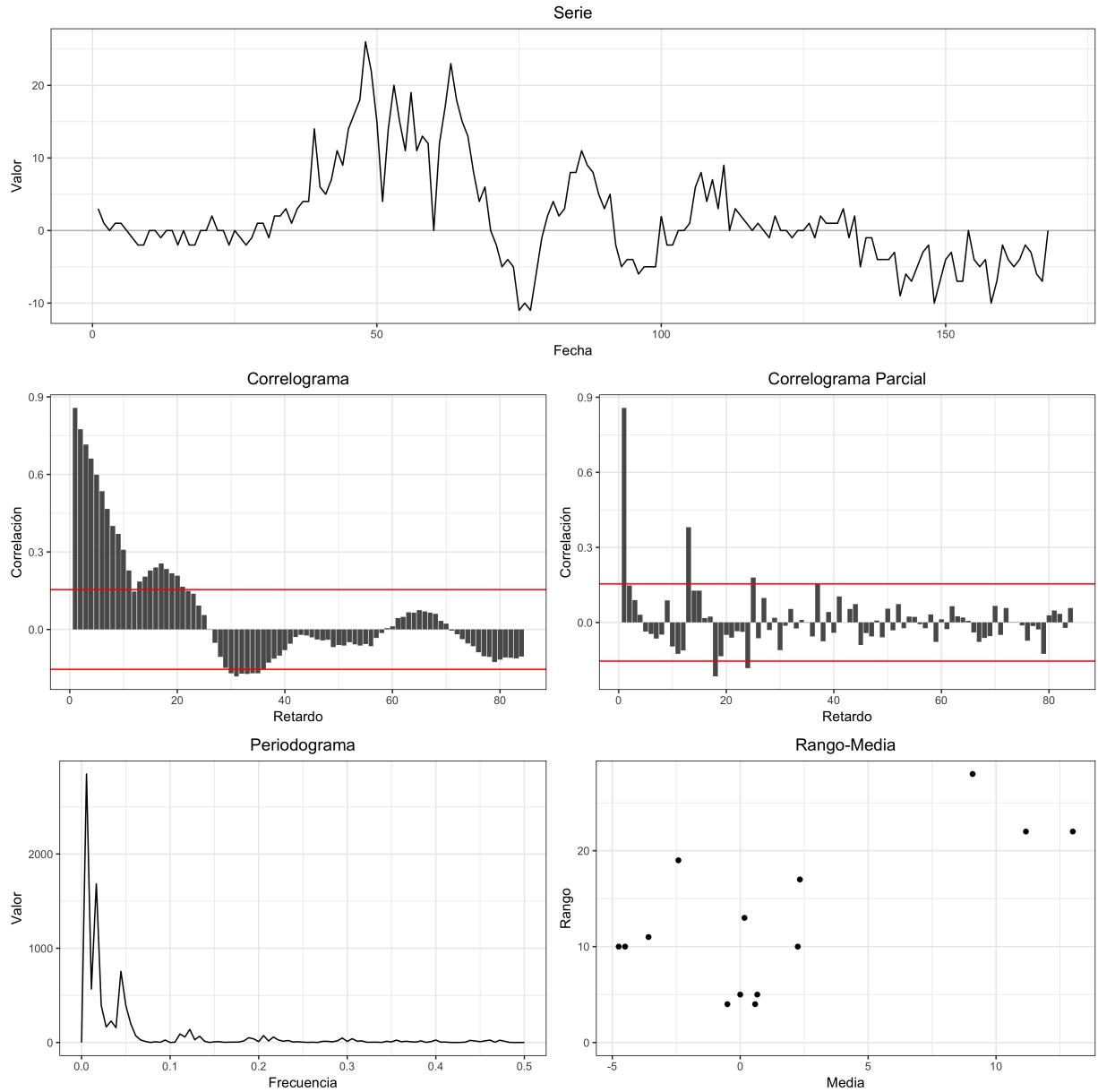


Figura 3: Gráfico de la serie, Correlograma, Correlograma Parcial, Periodograma y diagrama de dispersión *rango-media* para la serie `weightloss` tras la realización de una diferenciación estacional (12 retardos).

En el gráfico de la serie se puede apreciar la eliminación de la estructura estacional tras la serie, pero no la tendencia de la misma. Ahora la serie oscila en torno a valores próximos a cero, pero la tendencia hace que este comportamiento no pueda ser considerado estacionario. En el correlograma sucede algo similar: en este caso se ha conseguido eliminar la estructura lineal de decrecimientos en las correlaciones de la estacionalidad, pudiéndose considerar ahora decrecimiento exponencial. Sin embargo, la componente regular no tiene una estructura de decrecimientos exponenciales, lo cual hacer rechazar la hipótesis de que la serie sea estacionaria. En cuanto al correlograma parcial, se puede apreciar que se han reducido drásticamente las correlaciones referidas a los múltiplos de la estacionalidad, pero siguen destacando en gran medida las correlaciones en las posiciones 1 y 13. Se cree que la primera correlación se debe a la componente de tendencia de la serie, siendo la del 13-ésimo retardo un reflejo estacional de esta. En cuanto al periodograma, se puede apreciar que el primer armónico destaca en gran medida sobre el resto, lo cual se cree que es debido a la tendencia de la serie. En cuanto al diagrama de dispersión *rango-media*, se puede ver que la relación entre nivel y dispersión ha aumentado, lo cual es otro reflejo de que la serie resultante no es estacionaria.

La diferenciación estacional ha conseguido reducir en gran medida la componente estacional de la serie. En la Tabla 1 se indica que la varianza es 49,09, lo cual es una drástica reducción en la dispersión de la serie. Sin embargo ha dejado inalterada la tendencia de la misma, por lo que no se ha conseguido llegar a una serie estacionaria, que tomar de partida para el ajuste de un modelo autoregresivo con media móvil. Como se verá a continuación, la serie se transforma en estacionaria tras aplicar una combinación de las dos diferenciaciones realizadas previamente. Esto es, la diferenciación regular y la diferenciación regular, que aisla tanto la tendencia de la serie, como la estructura estacional de la misma.

1.3.3. Diferenciación regular y estacional

Tras analizar los resultados obtenidos al aplicar la diferenciación regular y la diferenciación estacional, para la cual se producen distintas modificaciones de la serie original que la dejan más cerca de la propiedad de estacionariedad, el siguiente paso es combinar ambas estrategias. Dado que la operación de diferenciación es conmutativa, no importa el orden de aplicación. En este caso, se ha desarrollado la expresión conjunta en la Ecuación 3.

$$\begin{aligned}\nabla \nabla_{12} X_t &= (1 - B) \cdot (1 - B^{12}) \cdot X_t \\ &= (1 - B) \cdot (X_t - X_{t-12}) \\ &= X_t - X_{t-1} - X_{t-12} + X_{t-13}\end{aligned}\tag{3}$$

En la Figura 4 se muestran distintos resúmenes visuales de la serie diferenciada regular y estacionalmente. En el gráfico de la serie se observa como la tendencia de la serie ha sido eliminada, al igual que la estacionalidad, lo cual hace que el comportamiento de la serie sea mucho más similar a un ruido blanco, aunque sin llegar a serlo, tal y como se puede apreciar a partir del resto de gráficos. Llama la atención el aumento de la dispersión en torno al primer cuarto de la serie, lo cual asumimos que podría darse al cambio de nivel durante finales del año 2008 en la serie original. En cuanto al correlograma, se puede apreciar un decrecimiento exponencial en las primeras correlaciones. De la misma manera, también se puede apreciar un decrecimiento exponencial en las correlaciones estacionales. Estas son las de la forma $12i$, es decir, las correlaciones múltiplo de 12. Dichos decrecimientos exponenciales son un indicador de que la serie resultante será estacionaria. Sobre el correlograma parcial, en este caso también se muestran decrecimientos exponenciales. Sin embargo, tanto este fenómeno como un análisis más detallado del correlograma, serán llevados a cabo en la Subsección 1.4 cuando se propongan distintos modelos *SARIMA*. En el periodograma se puede apreciar como destacan de una serie de armónicos sobre el resto. Sin embargo, en este caso no lo hacen de manera determinista, por lo que no pueden ser considerados una única estacionalidad que poder aislar de manera sencilla. Por lo tanto, lo trataremos como una componente aleatoria. Por último, a través del diagrama de dispersión *rango-media* se puede comprobar que la relación entre nivel y dispersión ha desaparecido. Entonces, ya podemos considerar que la serie resultante cumple la propiedad de *homocedastecidad*.

Dado que la serie cumple la propiedad de media y varianza constantes a lo largo del tiempo y además presenta decrecimientos exponenciales en el correlograma, consideraremos que la serie resultante es estacionaria. Además, la varianza de 14,07 es la menor que se ha obtenido, tal y como se puede apreciar en la

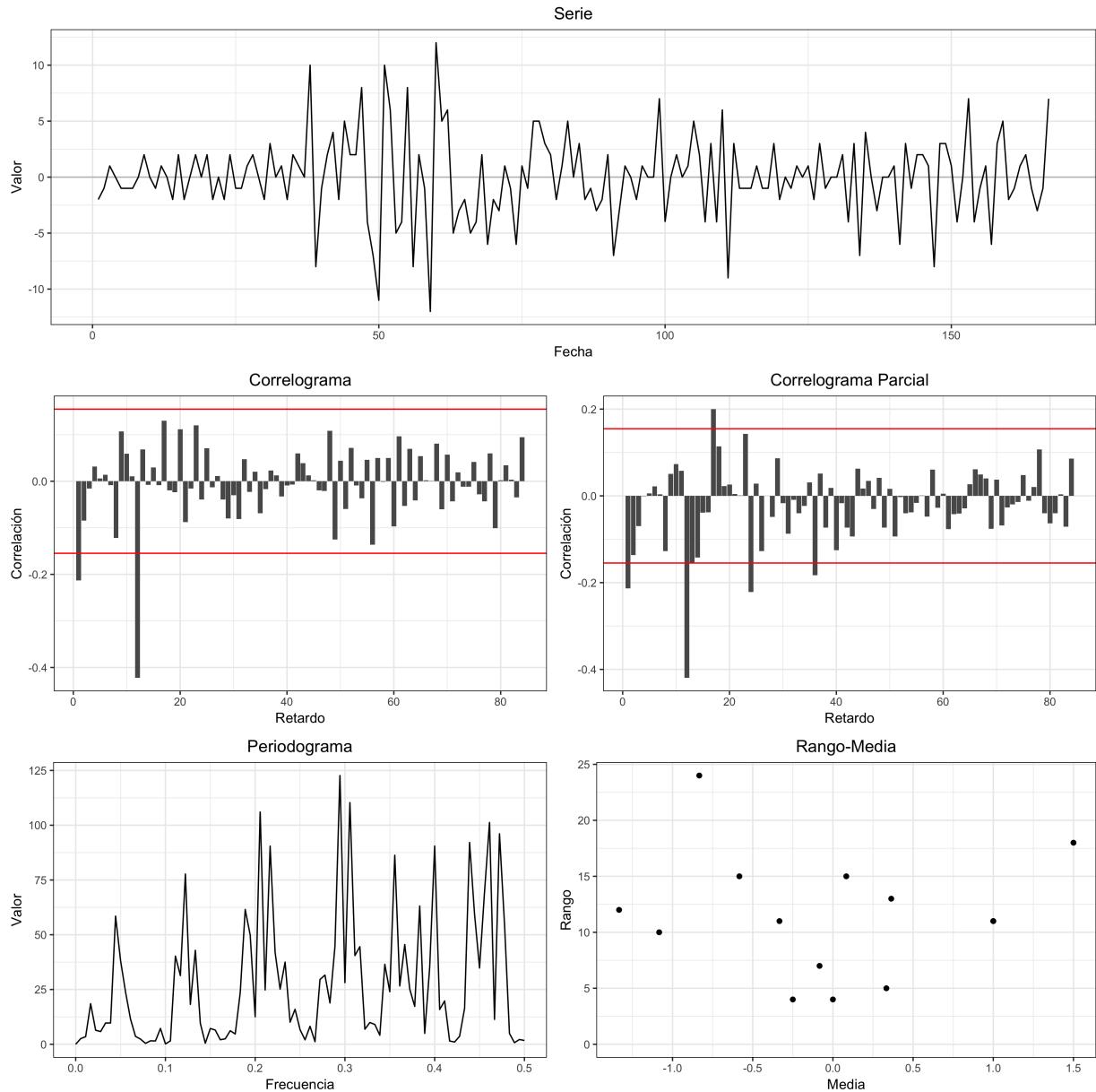


Figura 4: Gráfico de la serie, Correlograma, Correlograma Parcial, Periodograma y diagrama de dispersión *rango-media* para la serie `weightloss` tras la realización de una diferenciación regular y otra diferenciación estacional (12 retardos).

Tabla 1. Por estas razones, diremos que una diferenciación regular y otra estacional son suficientes para conseguir la propiedad de estacionaridad, necesaria para el ajuste de un modelo *SARIMA*.

Una vez realizado el análisis descriptivo de la serie, así como la búsqueda por los órdenes de diferenciación necesarios para conseguir estacionarizar nuestra serie, ya estamos en condiciones de analizar la serie desde otro punto de vista: la proposición de modelos, lo cual se expone a continuación.

1.4. Modelos propuestos

Tal y como se ha dicho posteriormente, en este apartado se van a proponer distintos modelos para el ajuste de la serie. Estos se proponen teniendo en cuenta únicamente el análisis descriptivo de la serie original, así como de las diferenciadas. Puesto que ya se ha decidido el grado de diferenciación de la serie en la Subsubsección 1.3.2 este será un valor fijo en la proposición de modelos. Esto se debe a que el ajuste de este tipo de modelos requiere de la estacionarización de la serie original.

Para indicar los modelos se utilizará la notación que se describe en la Ecuación 4, donde la primera tupla se refiere a los parámetros regulares, mientras que las subsiguientes se refieren al periodo estacional s_i -ésimo en cada caso. El parámetro p indica el orden de la componente autoregresiva en cada tupla, el parámetro d se refiere al grado de diferenciación y por último, el parámetro q al orden de la componente de media móvil.

$$\text{SARIMA}(p, d, q)(P_1, D_1, Q_1)_{s_1}(P_2, D_2, Q_2)_{s_2} \dots (P_q, D_q, Q_q)_{s_q} \quad (4)$$

Puesto que anteriormente se indicó que era necesaria una diferenciación regular y otra estacional (de longitud 12) en la Subsubsección 1.3.3, sabemos que los modelos propuestos serán de la forma $\text{SARIMA}(p, 1, q)(P, 1, Q)_{12} \dots (P_q, D_q, Q_q)_{s_q}$. Para proponer modelos, nos vamos a fijar en la Figura 4, que recoge distintos gráficos sobre la serie diferenciada regular y estacionalmente. En concreto, nos fijaremos sobre todo tanto en el correlograma como en el correlograma parcial, el cual sirve de guía a la hora de elegir órdenes de las componentes autorregresivas y de medias móviles.

El primer modelo que se propone es el indicado en la Ecuación 5, el cual se basa en la utilización de 1 parámetro de media móvil en la parte regular y otro en la parte estacional de periodicidad 12. La justificación detrás de este modelo es la siguiente: se pueden apreciar decrecimientos exponenciales en el diagrama de correlación parcial que comienzan en los retardos 1 y 12. Además, se dan valores elevados en el correlograma en las correlaciones 1 y 12, lo cual implica que un modelo de esta forma podría ajustarse de manera coherente a la serie.

$$\text{SARIMA}(0, 1, 1)(0, 1, 1)_{12} \quad (5)$$

Se intuye que el modelo de la Ecuación 5 será el más adecuado para la serie **weightloss**. Sin embargo, a continuación se proponen otros modelos para los que se cree que existen distintas razones por las cuales se deberían comportar de manera adecuada.

El modelo de la Ecuación 6 consiste en una ampliación respecto del anterior. En este se ha añadido 1 parámetro de media móvil más, en el retardo 17. La razón por la cual se propone este modelo es el decrecimiento exponencial en el correlograma parcial de la Figura 4, que comienza en dicho retardo.

$$\text{SARIMA}(0, 1, 1)(0, 1, 1)_{12}(0, 0, 1)_{17} \quad (6)$$

Para completar la proposición de modelos que podrían ajustarse de manera adecuada a la serie, se añaden los indicados en la Ecuación 7 y en la Ecuación 8, los cuales son ampliaciones de los indicados en la Ecuación 5 y en la Ecuación 6 respectivamente. Estos añaden 1 parámetro autoregresivo a la componente regular de la serie. La justificación sobre este parámetro se debe al decrecimiento exponencial a partir del segundo retardo en el correlograma (ya que el valor del primer retardo proviene de la componente de media móvil). Si estos modelos finalmente son considerados válidos, entonces también habrá que considerar la primera correlación parcial como procedente de la componente autoregresiva regular. Sin embargo, se intuye que estos modelos finalmente serán invalidados.

$$\text{SARIMA}(1, 1, 1)(0, 1, 1)_{12} \quad (7)$$

$$\text{SARIMA}(1, 1, 1)(0, 1, 1)_{12}(0, 0, 1)_{17} \quad (8)$$

El siguiente paso es proceder al ajuste de los modelos propuestos, prestando especial atención a la significancia de los tests para los parámetros en cada modelo, así como a los residuos obtenidos, los cuales permitirán validar (e invalidar) aquellos modelos que no se ajusten de manera satisfactoria a los datos.

2. Etapa de estimación y validación

Una vez propuestos distintos modelos que podrían ajustar de manera adecuada a la serie, a continuación se procederá al ajuste de los mismos, para ello se utilizará el lenguaje *SAS*.

2.1. Estimación y Validación de SARIMA(0, 1, 1)(0, 1, 1)₁₂

[TODO]

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.33162	0.07282	4.55	<.0001	1
MA2,1	0.62239	0.06576	9.46	<.0001	12

Tabla 2: [TODO].

[TODO]

Correlations of Parameter Estimates		
Parameter	MA1,1	MA2,1
MA1,1	1.000	0.037
MA2,1	0.037	1.000

Tabla 3: [TODO].

[TODO]

Variance Estimate	9.253677
Std Error Estimate	3.041986
AIC	853.4882
SBC	859.7242
Number of Residuals	167

Tabla 4: [TODO].

[TODO]

[TODO]

[TODO]

2.2. Estimación y Validación de SARIMA(0, 1, 1)(0, 1, 1)₁₂(0, 0, 1)₁₇

[TODO]

[TODO]

[TODO]

[TODO]

[TODO]

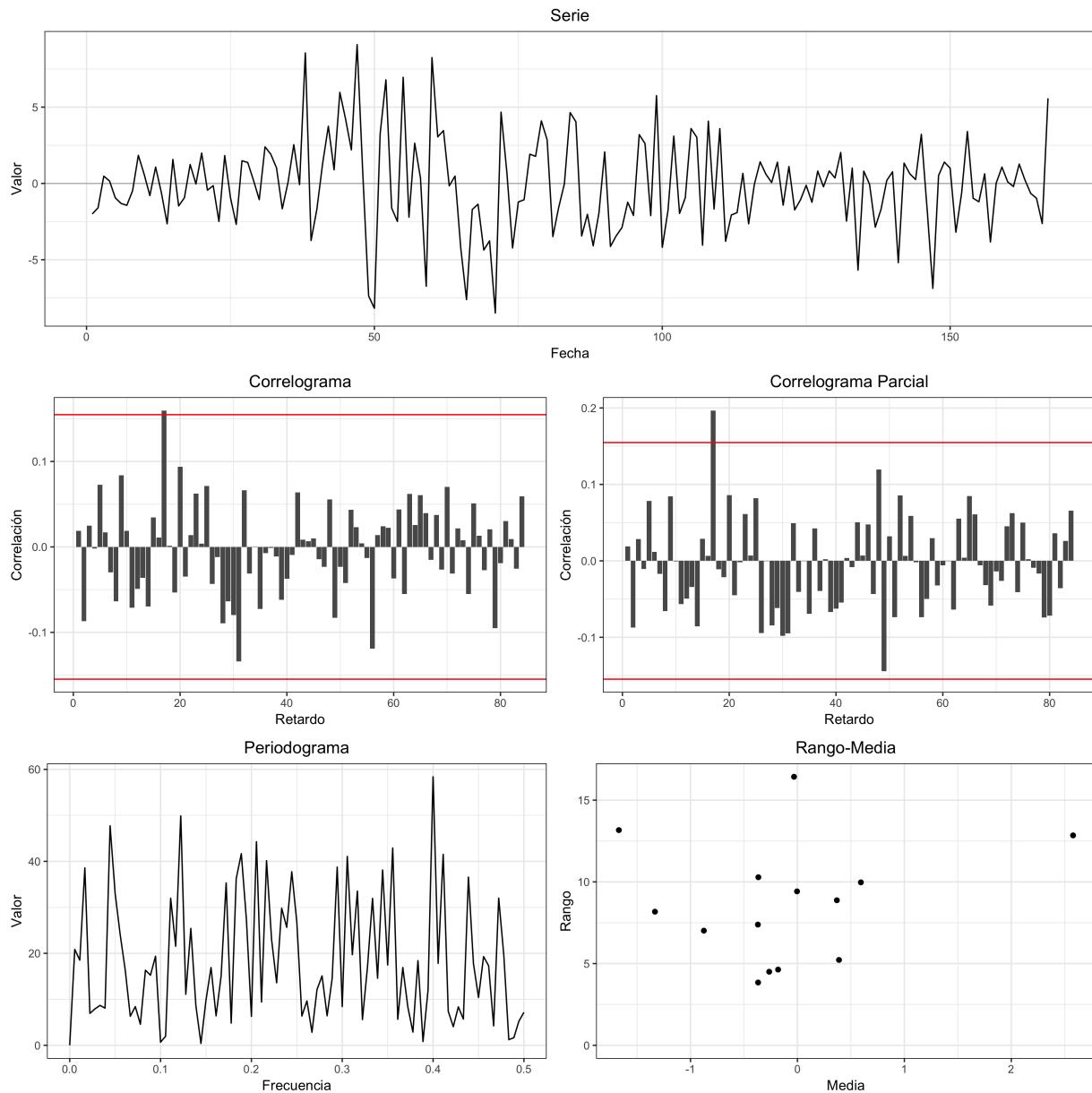


Tabla 5: [TODO].

[TODO]

2.3. Estimación y Validación de SARIMA(1, 1, 1)(0, 1, 1)₁₂

[TODO]

[TODO]

[TODO]

[TODO]

2.4. Estimación y Validación de SARIMA(1, 1, 1)(0, 1, 1)₁₂(0, 0, 1)₁₇

[TODO]

[TODO]

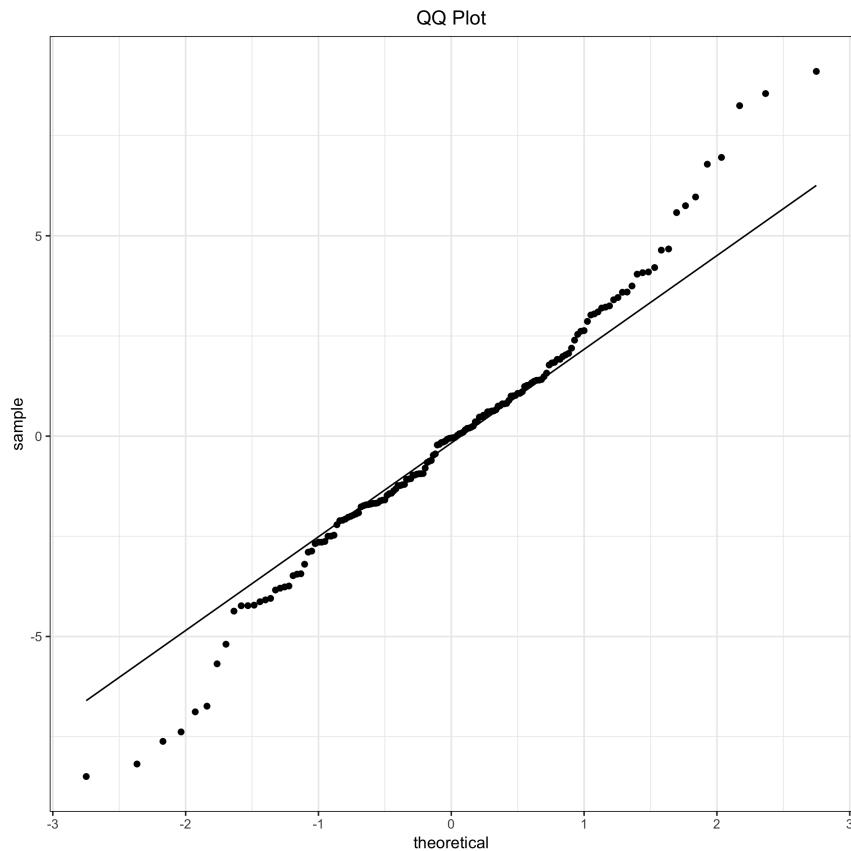


Tabla 6: [TODO].

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.33629	0.07278	4.62	<.0001	1
MA2,1	0.62106	0.06586	9.43	<.0001	12
MA3,1	-0.15874	0.07817	-2.03	0.0423	17

Tabla 7: [TODO].

Correlations of Parameter Estimates			
Parameter	MA1,1	MA2,1	MA3,1
MA1,1	1.000	0.027	-0.025
MA2,1	0.027	1.000	-0.021
MA3,1	-0.025	-0.021	1.000

Tabla 8: [TODO].

[TODO]

[TODO]

Variance Estimate	9.047408
Std Error Estimate	3.007891
AIC	851.1135
SBC	860.4674
Number of Residuals	167

Tabla 9: [TODO].

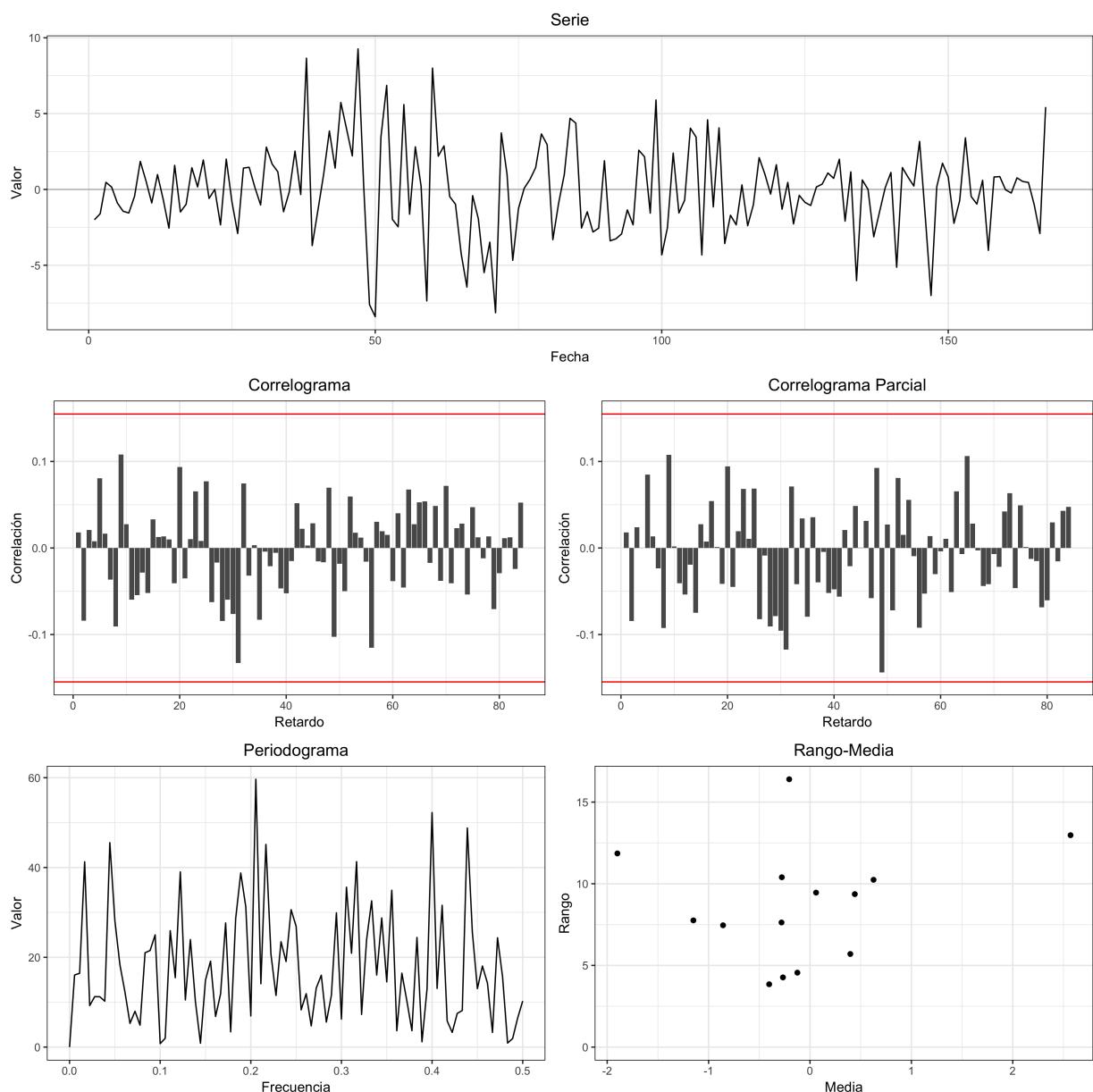


Tabla 10: [TODO].

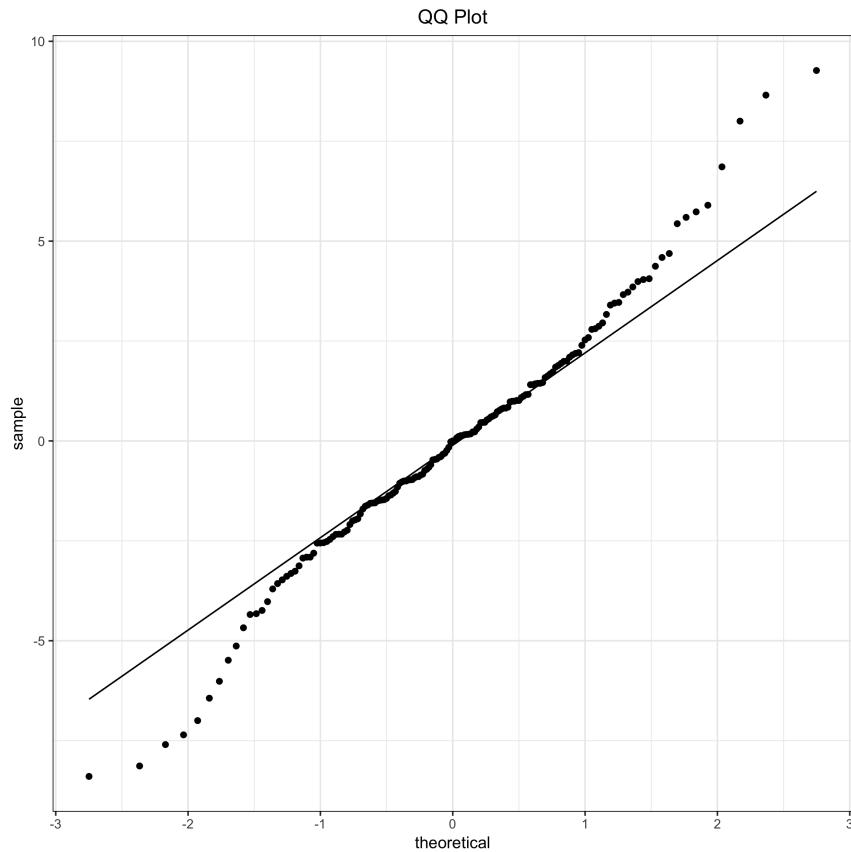


Tabla 11: [TODO].

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.46545	0.20516	2.27	0.0233	1
MA2,1	0.62276	0.06616	9.41	<.0001	12
AR1,1	0.15694	0.22980	0.68	0.4947	1

Tabla 12: [TODO].

Correlations of Parameter Estimates			
Parameter	MA1,1	MA2,1	AR1,1
MA1,1	1.000	-0.058	0.943
MA2,1	-0.058	1.000	-0.072
AR1,1	0.943	-0.072	1.000

Tabla 13: [TODO].

Variance Estimate	9.273561
Std Error Estimate	3.045252
AIC	854.8415
SBC	864.1955
Number of Residuals	167

Tabla 14: [TODO].

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.46122	0.20376	2.26	0.0236	1
MA2,1	0.62117	0.06629	9.37	<.0001	12
MA3,1	-0.15758	0.07845	-2.01	0.0446	17
AR1,1	0.14754	0.22799	0.65	0.5175	1

Tabla 15: [TODO].

Correlations of Parameter Estimates				
Parameter	MA1,1	MA2,1	MA3,1	AR1,1
MA1,1	1.000	-0.067	0.017	0.941
MA2,1	-0.067	1.000	-0.024	-0.078
MA3,1	0.017	-0.024	1.000	0.025
AR1,1	0.941	-0.078	0.025	1.000

Tabla 16: [TODO].

Variance Estimate	9.071058
Std Error Estimate	3.01182
AIC	852.5246
SBC	864.9965
Number of Residuals	167

Tabla 17: [TODO].

3. Comparación de modelos

[TODO]

Modelo	SSE_p
SARIMA(0, 1, 1)(0, 1, 1) ₁₂	49.84
SARIMA(0, 1, 1)(0, 1, 1) ₁₂ (0, 0, 1) ₁₇	42.73

Tabla 18: [TODO].

4. Predicción

[TODO]

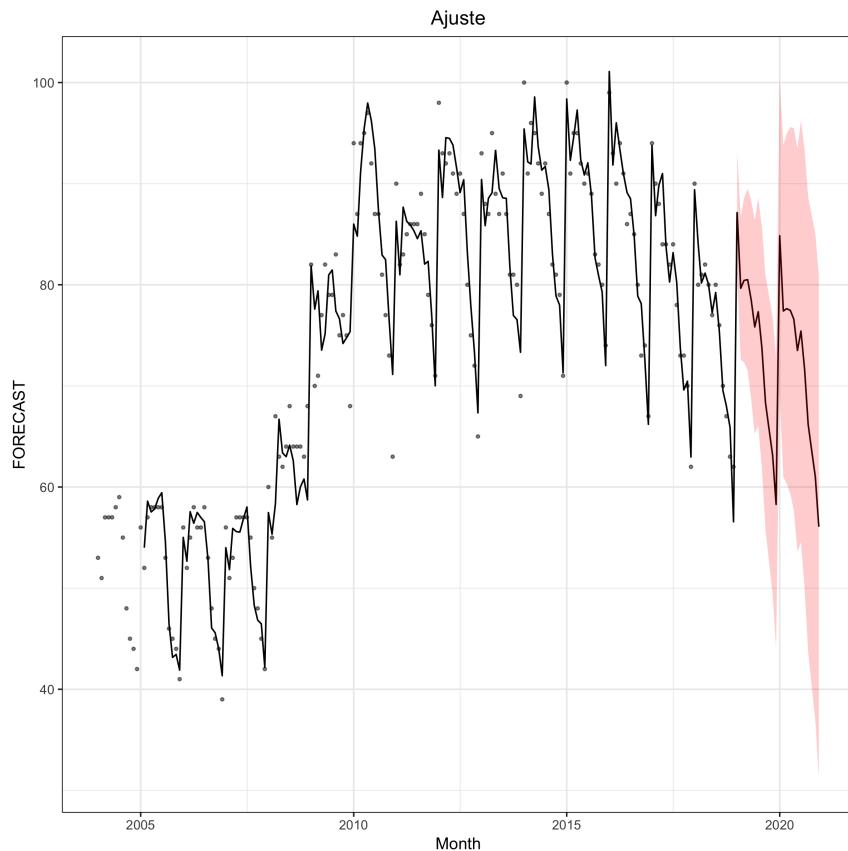


Tabla 19: [TODO].

[TODO]

[TODO]

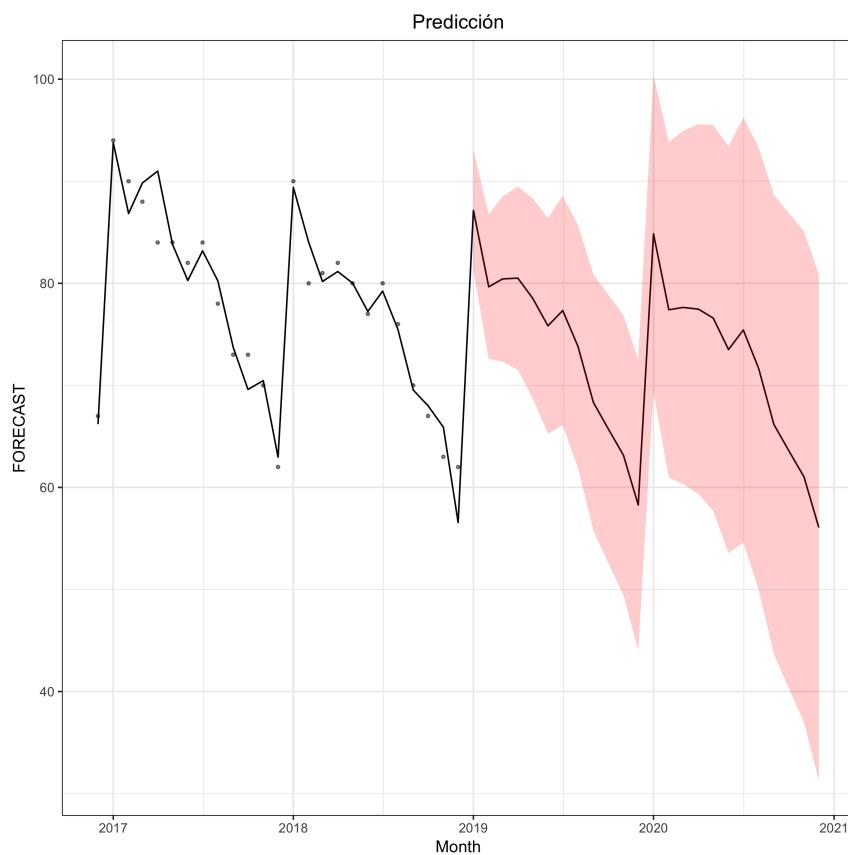


Tabla 20: [TODO].

A. Código Fuente

```

RangeMean <- function(x, seasonality) {
  n <- length(x)
  seq(1, n, by=seasonality) %>%
    sapply(function(i){
      a <- x[i:(i + seasonality - 1)]
      c(mean=mean(a, na.rm=TRUE), range=diff(range(a, na.rm = TRUE)))
    }) %>%
    t() %>%
    as.data.frame()
}

Correlogram <- function(x, n = length(x) - 1) {
  result <- acf(x, lag.max=n, plot=FALSE)$acf[1:n + 1]
  data.frame(lag = 1:length(result), values = result)
}

PartialCorrelogram <- function(x, n = length(x) - 1) {
  result <- pacf(x, lag.max=n, plot=FALSE)$acf
  data.frame(lag = 1:length(result), values = result)
}

Periodogram <- function(x) {
  result <- TSA:::periodogram(x, plot=FALSE)
  data.frame(freq = c(0, result$freq), spec = c(0, result$spec))
}

PredictionError <- function(df, lags) {
  sum(df[(nrow(df) - lags + 1):nrow(df), 'RESIDUAL'] ^ 2)
}

```

Código Fuente 1: [TODO]

```

source("res/code/functions.r")

library(ggplot2)

PlotTimeSeries <- function(df, seasonality, armonics = c(), lags = MAX_LAG){
  p.a <- ggplot(df) +
    aes(x = index, y = values) +
    xlab("Fecha") +
    ylab("Valor") +
    geom_hline(yintercept = 0, color = "gray") +
    geom_line() +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggttitle('Serie')

  p.b <- ggplot(RangeMean(df$values, seasonality)) +
    aes(x = mean, y = range) +
    geom_point() +
    xlab("Media") +
    ylab("Rango") +
    expand_limits(y=0) +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggttitle('Rango-Media')

  p.c <- ggplot(Correlogram(df$values, lags)) +
    aes(x = lag, y = values) +
    xlab("Retardo") +
    ylab("Correlación") +
    geom_bar(stat="identity") +
    geom_hline(yintercept = 2/sqrt(nrow(df)), color = "red") +
    geom_hline(yintercept = -2/sqrt(nrow(df)), color = "red") +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggttitle('Correlograma')

  p.partial.correlogram <- ggplot(PartialCorrelogram(df$values, lags)) +
    aes(x = lag, y = values) +
    xlab("Retardo") +
    ylab("Correlación") +
    geom_bar(stat="identity") +
    geom_hline(yintercept = 2/sqrt(nrow(df)), color = "red") +
    geom_hline(yintercept = -2/sqrt(nrow(df)), color = "red") +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggttitle('Correlograma Pcial')

  p.d <- ggplot(Periodogram(df$values)) +
    aes(x = freq, y = spec) +
    xlab("Frecuencia") +
    ylab("Valor") +
    geom_line() +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5),
          panel.border = element_rect(colour = "black", fill=NA)) +
    ggttitle('Periodograma')

  for (a in armonics) {
    p.d <- p.d + geom_vline(xintercept = a, color = "red", alpha = 0.4)
  }
  plot_grid(p.a, plot_grid(p.c, p.partial.correlogram, p.d, p.b, ncol = 2),
            ncol = 1, rel_heights = c(1, 2))
}

```

Código Fuente 2: [TODO]

```

rm(list = ls())

library(magrittr)
library(dplyr)
library(latex2exp)
require(reshape2)
library(forecast)
library(cowplot)
library(lubridate)

source("res/code/functions.r")
source("res/code/plotting.r")

BASE_PATH <- './'
BASE_IMG_PATH <- paste0(BASE_PATH, 'res/img/')
BASE_DATA_PATH <- paste0(BASE_PATH, 'res/data/')

weightloss <- read.csv(paste0(BASE_DATA_PATH, 'weight-loss.csv'))
weightloss$Month <- ymd(weightloss$Month, truncated = 2)
colnames(weightloss) <- c("index", "values")

PlotTimeSeries(weightloss, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

values <- diff(weightloss$values, 1)
df <- data.frame(index = 1:length(values), values=values)
PlotTimeSeries(df, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss-diff-1.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

values <- diff(weightloss$values, 12)
df <- data.frame(index = 1:length(values), values=values)
PlotTimeSeries(df, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss-diff-12.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

values <- diff(weightloss$values, 12, 2)
df <- data.frame(index = 1:length(values), values=values)
PlotTimeSeries(df, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss-diff-12-12.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

values <- diff(diff(weightloss$values, 1), 12)
df <- data.frame(index = 1:length(values), values=values)
PlotTimeSeries(df, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss-diff-1-12.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

values <- diff(diff(weightloss$values, 1), 12, 2)
df <- data.frame(index = 1:length(values), values=values)
PlotTimeSeries(df, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss-diff-1-12-12.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

values <- diff(diff(weightloss$values, 1, 2), 12)
df <- data.frame(index = 1:length(values), values=values)
PlotTimeSeries(df, seasonality = 12, lags = 84) %>%
  { save_plot(paste0(BASE_IMG_PATH, 'weightloss-diff-1-1-12.png')), ,
    base_aspect_ratio = 1, base_height = 12 }

```

Código Fuente 3: [TODO]

```

FILENAME REFFILE '/folders/myshortcuts/sarima-weight-loss/res/data/weight-loss.csv';

PROC IMPORT DATAFILE=REFFILE
DBMS=CSV
OUT=WORK.weightloss;
GETNAMES=YES;
RUN;

```

Código Fuente 4: [TODO]

```

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate q=(1)(12) noint method=ml;
  forecast id=month interval=month lead=0 out=arimaout_1;
run;

proc export data=arimaout_1
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/validation-1.csv'
  dbms=csv replace;
run;

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate q=(1)(12)(17) noint method=ml;
  forecast id=month interval=month lead=0 out=arimaout_2;
run;

proc export data=arimaout_2
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/validation-2.csv'
  dbms=csv replace;
run;

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate p=(1) q=(1)(12) noint method=ml;
  forecast id=month interval=month lead=0 out=arimaout_3;
run;

proc export data=arimaout_3
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/validation-3.csv'
  dbms=csv replace;
run;

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate p=(1) q=(1)(12)(17) noint method=ml;
  forecast id=month interval=month lead=0 out=arimaout_4;
run;

proc export data=arimaout_4
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/validation-4.csv'
  dbms=csv replace;
run;

```

Código Fuente 5: [TODO]

```

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate q=(1)(12) noint method=ml;
  forecast id=month interval=month lead=12 out=comparison_1 back = 12;
run;

proc export data=comparison_1
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/comparison-1.csv'
  dbms=csv replace;
run;

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate q=(1)(12)(17) noint method=ml;
  forecast id=month interval=month lead=12 out=comparison_2 back = 12;
run;

proc export data=comparison_2
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/comparison-2.csv'
  dbms=csv replace;
run;

```

Código Fuente 6: [TODO]

```

rm(list = ls())

source("res/code/functions.r")

BASE_PATH <- './'
BASE_DATA_PATH <- paste0(BASE_PATH, 'res/data/')

comparison.1 <- read.csv(paste0(BASE_DATA_PATH, 'comparison-1.csv'))
comparison.2 <- read.csv(paste0(BASE_DATA_PATH, 'comparison-2.csv'))

errors <- data.frame()
errors <- rbind(errors, list(name="${\\text{SARIMA}}(0, 1, 1)(0, 1, 1)_{12}$}",
                           error=PredictionError(comparison.1, 12)),
               stringsAsFactors = FALSE)
errors <- rbind(errors, list(name="${\\text{SARIMA}}(0, 1, 1)(0, 1, 1)_{12}(0, 0, 1)_{17}$}",
                           error=PredictionError(comparison.2, 12)),
               stringsAsFactors = FALSE)

errors$error <- round(errors$error, digits = 2)
write.csv(errors, paste0(BASE_DATA_PATH, 'predict-error.csv'), row.names = FALSE, quote = FALSE)

```

Código Fuente 7: [TODO]

```

proc arima data = weightloss;
  identify var=weight_loss(1, 12);
  estimate q=(1)(12)(17) noint method=ml;
  forecast id=month interval=month lead=24 out=predict;
run;

proc export data=predict
  outfile='/folders/myshortcuts/sarima-weight-loss/res/data/predict.csv'
  dbms=csv replace;
run;

```

Código Fuente 8: [TODO]