

Análisis de Series Temporales: Tarea 2 - Ejercicio 2

Alejandro Del Hierro Diez

Gabriel Rodríguez Canal

Miguel Martín Mateos

Sergio García Prado

3 de diciembre de 2018

- **Archivo:** `tuberculo.sf3`

- **Serie:** Número de casos registrados semanalmente de tuberculosis respiratoria en España, entre los años 1982 y 1991 (el primer dato corresponde al número de casos registrados desde el 1 de Enero de 1982 al 7 de Enero de 1982).

- $\{X_t\}$ Serie Original.
- $\{Y_t\}$ Serie del numero de casos en periodos de cuatro semanas sucesivos.

- (a) **Describir estas dos series ($\{Y_t\}$ puede crearse con el `proc expand` de *SAS*), indicando claramente para cada una de ellas qué frecuencias elegiríais a priori para ajustar un modelo con tendencia polinómica más ondas.**

En este trabajo se va a analizar una serie temporal unidimensional referida al número de casos registrados en España de problemas re tuberculosis respiratoria. La serie original se da en periodos acumulados de manera semanal, por lo que el número de observaciones anuales es variable de 52. Sin embargo, nótese que este valor puede variar sutilmente debido a los días concretos de comienzo y fin de cada año.

Otra de las dificultades de este análisis procede de la enorme amplitud del periodo de la serie, que complica el análisis estacional, así como el ajuste de la serie a nivel computacional drásticamente. Por esta razón, se va a proceder al estudio de la serie mediante agrupamiento de las observaciones en periodos de 4 semanas. Esto implica $52/4 = 13$ observaciones anuales, las cuales se pueden analizar de una manera más cómoda sin perder demasiada información.

Para el estudio de las series en cuestión (así como su ajuste) se utilizará el sistema *SAS*, para el cual se incluyen distintos fragmentos de código que permiten la reconstrucción del análisis. La mayoría de dichos bloques de código se han incluido en el apartado en cuestión del documento. Sin embargo, otros que no se han considerado lo suficientemente relevantes para el estudio se han incluido en el Apéndice (I.I). Para la generación de algunos gráficos, así como otras partes del análisis, se ha utilizado el lenguaje *R*, cuyo código fuente se incluye en el Apéndice (I.II) por la misma razón que el caso anterior.

El resto del apartado se divide de la siguiente manera: en el Subapartado (a.1) se estudia la serie sin agrupar, para posteriormente realizar el agrupamiento y analizar la serie resultante en Subapartado (a.2).

(a.1) Serie Semanal: $\{X_t\}$

Para el análisis de la serie semanal, a la cual denotaremos por $\{X_t\}$ de forma matemática, lo primero es cargarla en el sistema *SAS*, lo cual se puede llevar a cabo a través de la ejecución del Código Fuente 15. Lo primero en que nos fijamos es el número de observaciones de la serie, las cuales son 520 (52 observaciones y un total de 10 años). Otra de las cuestiones que hay que tener en cuenta para que el análisis de la serie sea realizado de manera adecuada es fijar la fecha de manera coherente. Si recurrimos al calendario, el 1 de Enero de 1982 fue *Domingo*, lo cual debemos indicar para que las fechas sean construidas de manera adecuada dado que en las observaciones frontera entre un distintos años, esto puede influir de manera perjudicial para el análisis.

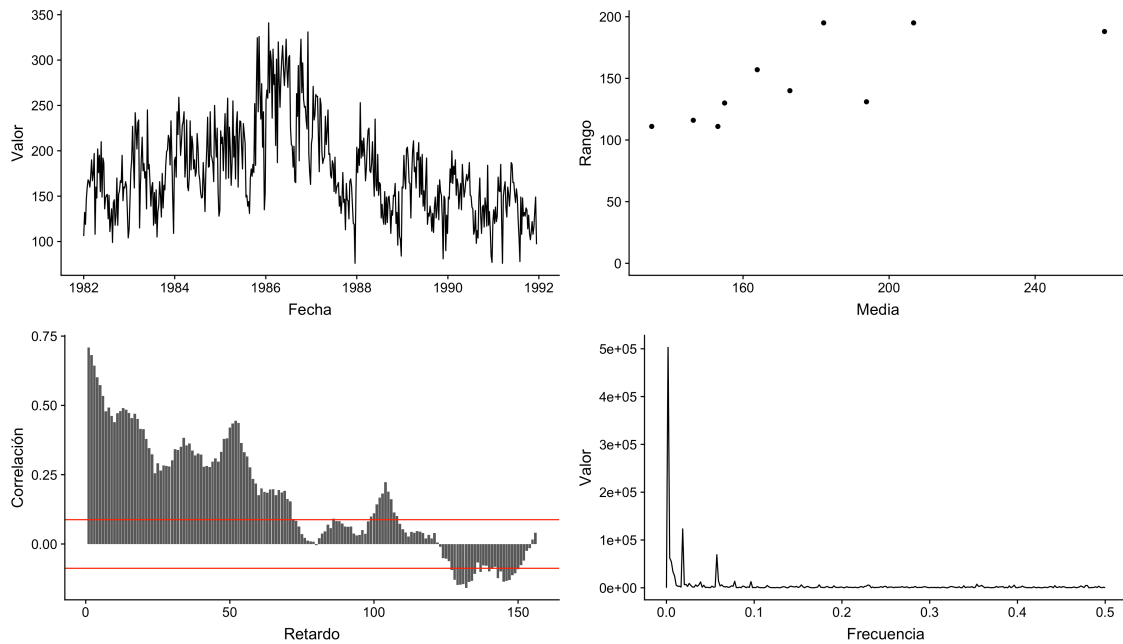


Figura 1: Gráfico de la serie, gráfico de dispersión *rango-media*, *correlograma* y *periodograma* del conjunto de datos *EJ2.SEMANAL*

Una vez cargada formateada de manera apropiada la fecha en la serie, ya se puede proceder al análisis descriptivo de la misma. Para ello se ha utilizado el Código Fuente 16, a través del cual hemos obtenido la Figura 1. Tal y como se puede apreciar a través del gráfico de la serie, esta presenta una tendencia que podría aproximarse a partir de un polinomio de orden cuadrático, alcanzando su máximo entre los años 1986 y 1987. Algo interesante es la forma de dicha tendencia, que experimenta un crecimiento de forma lineal desde 1981 hasta su máximo, para después sufrir un fuerte decrecimiento (que también se podría considerar lineal) durante el año 1987. El resto de la serie ([1988,1992]) mantiene una tendencia que podría ser considerada constante. Por lo tanto, esta serie presenta cambios de tendencia (lo cual requerirá modelos basados en *Suavizado Exponencial* tal y como veremos en los próximos apartados). Además de esto, el gráfico de la serie parece mostrar una componente estacional, sin embargo es difícil ver esto, por lo que posteriormente procederemos a diferenciarla para tratar de visualizarlo de manera más clara. Sin embargo, tal y como se indica a continuación, esto se aprecia mejor en la serie agrupada cada cuatro semanas $\{Y_t\}$.

Continuando con los diagramas de la Figura 1, a partir del gráfico de dispersión *rango-media*, se puede apreciar cierta relación entre estas, lo cual implica una relación entre el nivel y la dispersión en la serie. Es decir, cuanto mayor es el valor medio en la serie, mayor es la varianza de los datos. Para corregir esto existen distintos métodos, entre los que se encuentran transformaciones de tipo *Box-Cox* que tratan de eliminar dicha relación. Sin embargo, el estudio de dichos métodos no son el objetivo de este análisis.

En cuanto al *correlograma*, en este se aprecia una alta correlación entre las variables (lo cual indica tendencia). Estos decrecimientos parecen localmente exponenciales (lo cual indicaría estacionaridad local), por lo que modelos basados en suavizado exponencial deberían ajustarse bien a la serie. A partir de este gráficos, podemos intuir la estructura estacional de la serie. Sin embargo, es complicado poder estudiarla detalladamente debido a la tendencia, así como el enorme periodo de la serie. Aunque destaca sobre el resto la correlación 52 como era de esperar, ya que indica el periodo de la serie. También son llamativas las próximas a las correlaciones 17 y 34, lo cual podría indicar un periodo *cuatrimestral*.

Para finalizar, en el *periodograma* se aprecia que el primer armónico destaca fuertemente sobre el resto, lo cual se debe a la tendencia de la serie. Debido al elevado valor que toma dicho armónico, no es posible estudiar el resto de valores de manera sencilla.

Para tratar de reducir la componente de tendencia y poder analizar de manera más cómoda la estacionalidad, procedemos a realizar una diferenciación sobre la serie. En la Figura 2 se muestran los gráficos descriptivos referidos a la serie diferenciada.

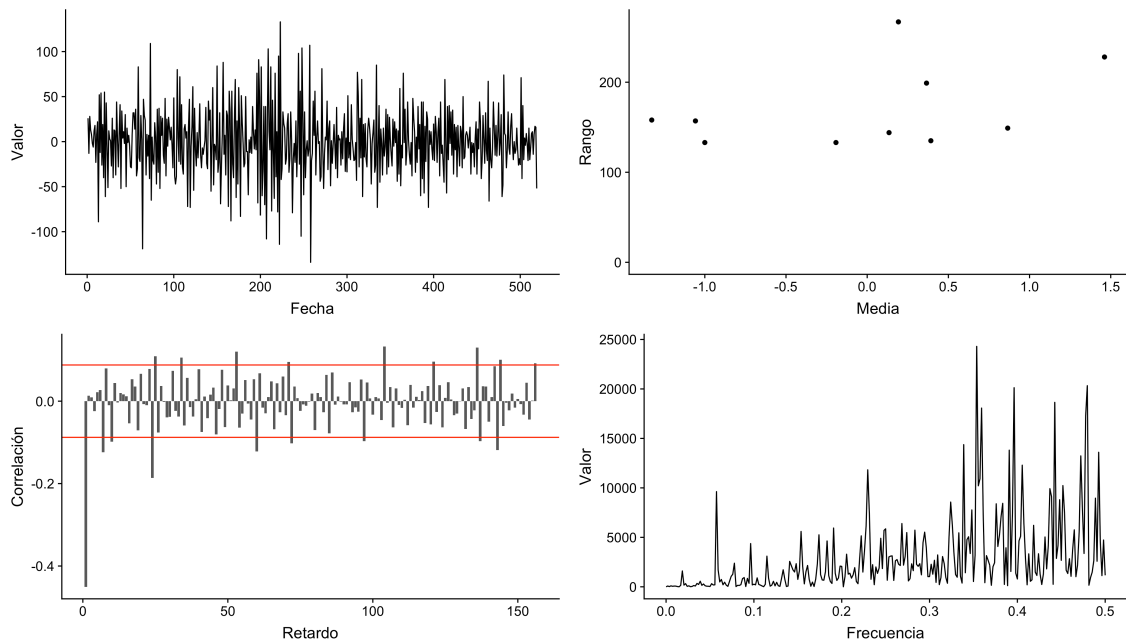


Figura 2: Gráfico de la serie, gráfico de dispersión *rango-media*, *correlograma* y *periodograma* del conjunto de datos EJ2.SEMANAL tras aplicar una diferenciación de primer orden.

Lo primero que llama la atención tras la diferenciación es la eliminación completa de la tendencia. En las zonas de crecimiento y decrecimiento ahora se muestra un mayor grado de variabilidad. También se aprecia en mayor medida la estructura estacional, pero nuevamente el enorme periodo de la serie no permite un análisis sencillo sobre la misma (en el Subapartado (a.2)) se puede estudiar más cómodamente la estacionalidad.

Siguiendo con la interpretación de los gráficos de la 2, la relación entre nivel y dispersión se ha reducido drásticamente, lo cual se puede comprobar a través del gráfico *rango-media*. En cuanto al *correlograma*, este muestra una alta correlación en sentido negativo con el primer retardo. Sin embargo, lo que más destaca es la estructura cíclica del resto de correlaciones, la cual se puede apreciar mucho mejor en la serie agrupada que comentaremos posteriormente. Para finalizar, en el *periodograma* se puede apreciar que hay armónicos que destacan sobre el resto y marcan tendencia, pero que estudiaremos en la serie agrupada cada cuatro semanas por comodidad y sencillez en el análisis.

(a.2) Serie 4 semanas: $\{Y_t\}$

Para trabajar con la serie agrupada en periodos de cuatro semanas. Es decir, la suma de observaciones de 4 en 4 se ha utilizado el Código Fuente 1. Este se basa en la sentencia `proc expand` que facilita el agrupamiento/separación de observaciones indexadas sobre un soporte temporal.

```
proc expand data = EJ2.SEMANAL out = EJ2.SEMANAL4 from = week.6 to = week4.6 ;
  id DATE;
  convert TUBERSEM / observed = total;
run;
```

Código Fuente 1: Generación del conjunto de datos EJ2.SEMANAL4 a partir de EJ2.SEMANAL

Una vez creado el conjunto de datos que se utilizará para el análisis, el siguiente paso es realizar un estudio descriptivo sobre el mismo. En este caso estamos trabajando con $520/4 = 130$ observaciones en periodos de 4 semanas, comenzando el 1 de Enero de 1982 que, como indicamos anteriormente, fue *Domingo*. Es de especial importancia darse cuenta de que los periodos de 4 semanas no coinciden con periodos mensuales. En este caso la amplitud del periodo es 13, mientras que en los análisis mensuales, es de 12 observaciones. Es necesario tener esto presente a lo largo del documento.

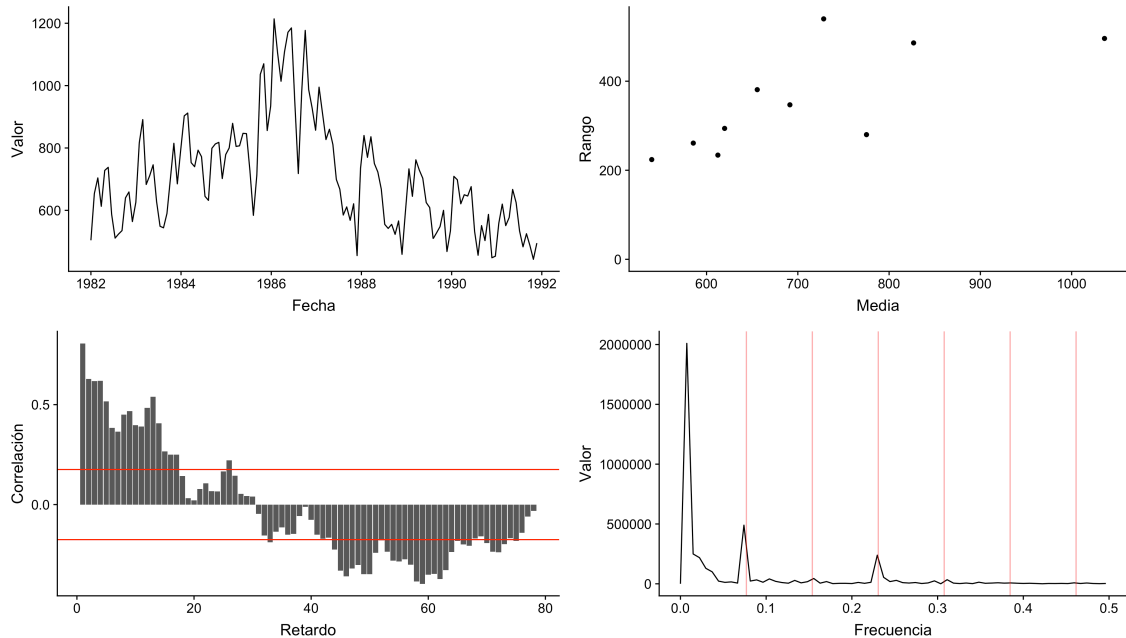


Figura 3: Gráfico de la serie, gráfico de dispersión *rango-media*, *correlograma* y *periodograma* del conjunto de datos EJ2.SEMANAL4

En la Figura 3 se muestran distintos gráficos descriptivos referidos a la serie agrupada cada 4 semanas $\{Y_t\}$, los cuales procedemos a analizar e interpretar a continuación. Lo más destacado (y obvio) de estos es que son un reflejo simplificado de los de la serie semanal $\{X_t\}$ que se muestran en la Figura 1.

Si nos fijamos en el gráfico de la serie, seguimos viendo la tendencia, que indicamos anteriormente, pero en este caso de una manera mucho más sencilla de interpretar. Ahora se ve más clara la estructura cuadrática de la misma, junto con una relación entre nivel y dispersión (que se confirma sobre el gráfico de dispersión *rango-media*). Otra de las partes que más llama la atención al visualizar el gráfico de la serie es que ahora es mucho más sencillo fijarse en la componente estacional de la serie, que sigue una estructura periodica tras cada 13 observaciones.

A partir del *correlograma* se puede apreciar la misma estructura de correlaciones que se indicaba sobre la serie semanal, donde destaca una estructura periódica (que estudiaremos más en detalle sobre la serie diferenciada). Al igual que antes, la serie no puede ser considerada estacionaria porque el decrecimiento entre correlaciones no es exponencial, aunque de la misma manera que antes, si que presenta esta estructura de manera local, lo cual nos confirma que un ajuste basado en suavizado exponencial debería ser acertado (sin olvidar la componente estacional).

En cuanto al *periodograma*, al igual que antes, uno de los primeros armónicos destacan fuertemente sobre el resto, lo cual intuimos que se debe a la tendencia de la serie. Además, en este caso se puede apreciar algo que no veíamos en la serie sin agrupar. Esto es que los armónicos de periodo $i/13$ destacan sutilmente sobre su entorno. En concreto, destacan sobre el resto los armónicos $1/13$ y $3/13$. Aunque el primero de ellos no parece hacerlo de manera determinista, ya que no coincide exactamente con el valor $1/13$, sino que es algo menor.

Para poder estudiar la componente estacional con mayor precisión se ha procedido a la diferenciación de la serie agrupada $\{Y_t\}$, cuyos gráficos descriptivos se muestran en la Figura 4. Al igual que antes, esta serie es una versión simplificada de la estudiada para la serie semanal $\{X_t\}$, que se muestra en la Figura 2.

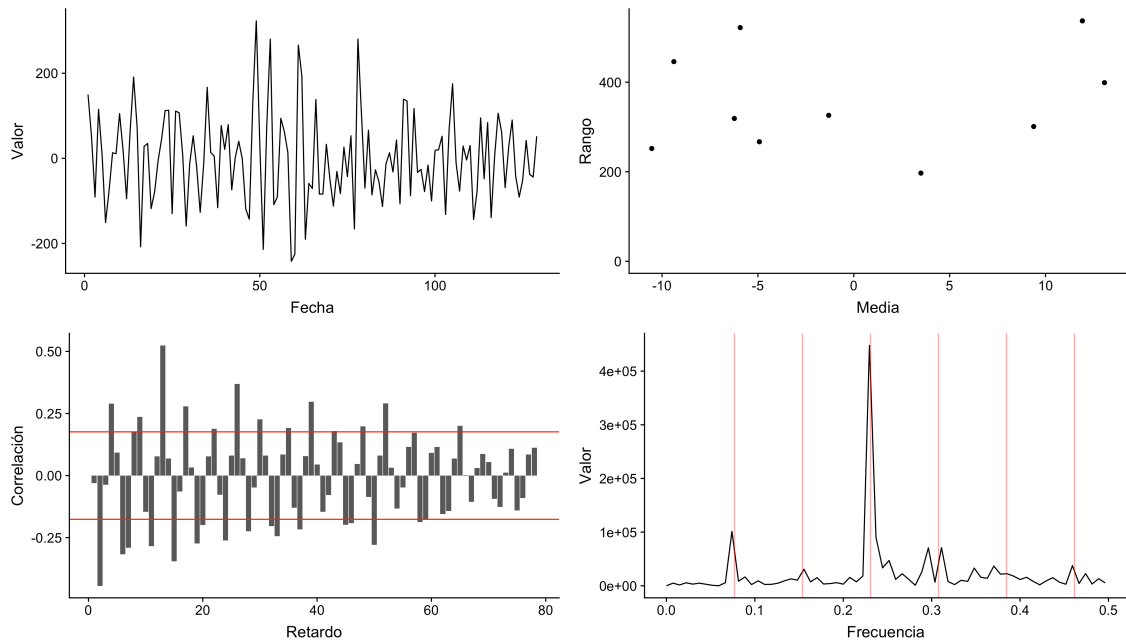


Figura 4: Gráfico de la serie, gráfico de dispersión *rango-media*, *correlograma* y *periodograma* del conjunto de datos EJ2.SEMANAL4 tras aplicar una diferenciación de primer orden.

El *gráfico de la serie* repite de manera cíclica el mismo patrón, lo cual confirma la existencia de estacionalidad, aunque a través de este diagrama es difícil de interpretar. Al igual que antes, el diagrama de dispersión *rango-media* muestra una estabilización de la dispersión, lo cual podría indicarnos que modelos basados en diferenciación no requerirán de una estabilización de la varianza previa mediante transformaciones. En cuanto al *correlograma*, ahora es trivial interpretar la existencia de estacionalidad, debido a la estructura de correlaciones, que se repite cíclicamente alternando entre valores positivos y negativos. En cuanto al *periodograma*, este nos muestra que los armónicos que siguen la forma $i/13$ destacan sobre el resto.

En el título del apartado se dice que es necesario indicar los armónicos significativos para la serie agrupada en periodos de 4 semanas y que denotamos por Y_t . Por tanto, tras el análisis de la misma, podemos confirmar que el **armónico referido a la frecuencia 3/13 es muy significativo**. También lo son, aunque de manera más sutil los armónicos 1/13, 4/13, 6/13 y 2/13 respectivamente. Aunque para confirmar su significancia se deberían los respectivos contrastes de hipótesis.

Antes de finalizar el apartado, se concluye que el agrupamiento de la serie semanal original en observaciones de 4 semanas es una estrategia que permite analizar la estructura inherente de la serie de manera mucho sencilla e intuitiva. Dado que el objetivo es tratar de comprender la estructura de la serie para así poder ajustar un modelo coherente para la misma, esto simplifica en gran medida el trabajo, no solo a nivel analítico, sino también a nivel computacional, dado que el número de parámetros a estimar se reduce drásticamente.

- (b) **Ajustar por suavizado exponencial, con el `proc esm`, los tres modelos que se consideren más apropiados para la serie $\{Y_t\}$ y comprobar su adecuación.**

Tras realizar un exhaustivo análisis de todas las diferentes posibilidades de ajuste por suavizado exponencial elegimos el *seasonal exponential smoothing*, el *suavizado de Winters multiplicativo* y el *aditivo*, ya que son los tres que incorporarán la estacionalidad de la serie.

(b.1) Suavizado Exponencial con Estacionalidad

```
proc esm data = EJ2.SEMANAL4 plot = (errors wn forecasts ACF) print = (all states);
  id DATE interval = WEEK4.6;
  forecast TUBERSEM / model = SEASONAL;
run;
```

Código Fuente 2: Código fuente para el ajuste de un modelo de *Suavizado Exponencial con Estacionalidad* sobre el conjunto de datos EJ2.SEMANAL4

Para comenzar explicaremos el modelo SES en el que mediante la tabla de los estimadores del parámetro y la significancia de dicho test.

Suavizado exponencial estacional Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.60356	0.05877	10.27	<.0001
Peso estacional	0.0010000	0.09043	0.01	0.9912

Tabla 1: Significancia para el modelo de *Suavizado Exponencial con Estacionalidad* sobre el conjunto de datos EJ2.SEMANAL4

Observamos que para la constante de suavizado estacional no se rechaza su significancia a cualquier nivel ya que el $p - valor$ del test se encuentra próximo 0,99.

En cambio, si se rechazará para la constante de suavizado para la media, donde $p - valor < 0,001$. Será conveniente diferenciar ya que el estimador estacional es cercano a 0. A continuación adjuntamos el gráfico del ACF de residuales:

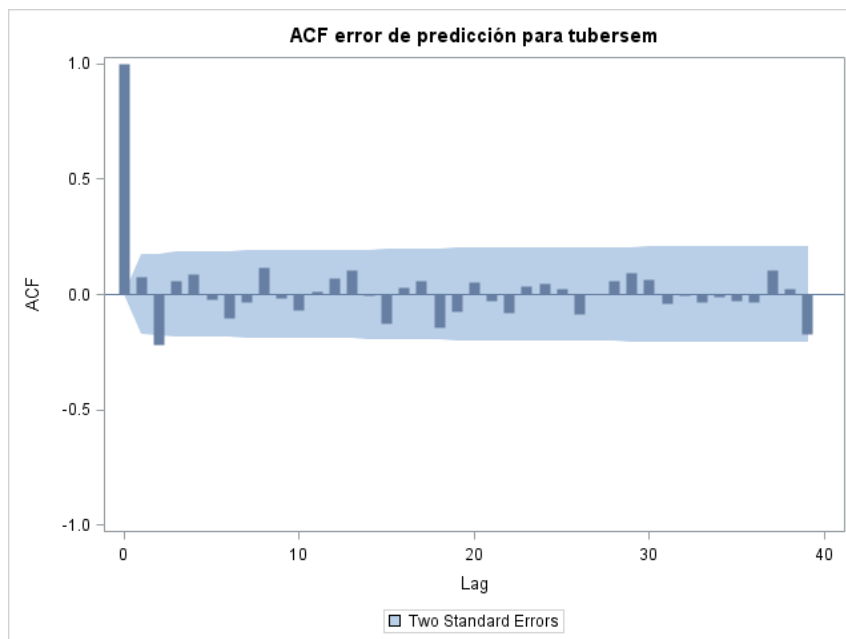


Figura 5: Gráfico de autocorrelaciones (correlograma) para los residuales del modelo de *Suavizado Exponencial con Estacionalidad* sobre el conjunto de datos EJ2.SEMANAL4

Observamos como dicho modelo ofrece dudas sobre su validación ya que para el retardo 2 la autocorrelación es muy alta. El retardo 1 no es muy alto, no resultará perjudicial para el modelo. Los retardos estacionales(cada 13) no son notables por lo que es bueno para la validación.

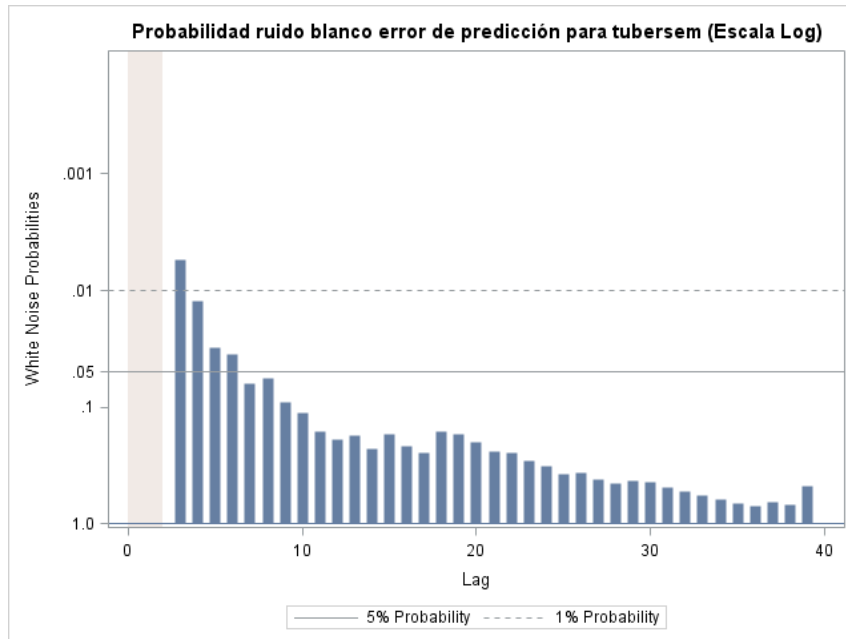


Figura 6: Gráfico sobre el contraste de ruido blanco para los residuales del modelo de *Suavizado Exponencial con Estacionalidad* sobre el conjunto de datos EJ2.SEMANAL4

Vemos en la gráfica que el test de que las correlaciones sean cero se rechaza para los 4 primeros retardos a nivel 0.05. Esto es indeseable para validar el modelo puesto que no podemos asegurar que sea ruido blanco, que es lo que se busca.

A continuación, pasamos a realizar el ajuste del modelo aditivo de Winter.

(b.2) Winter Aditivo

```
proc esm data = EJ2.SEMANAL4 plot = (errors wn forecasts ACF) print = (all states);
  id DATE interval = WEEK4.6;
  forecast TUBERSEM / model = ADDWINTERS;
run;
```

Código Fuente 3: Ajuste de un modelo de *Winter Aditivo* sobre el conjunto de datos EJ2.SEMANAL4

Método de Winters (Aditivo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.60191	0.05845	10.30	<.0001
Peso de tendencia	0.0010000	0.0071382	0.14	0.8888
Peso estacional	0.0010000	0.07877	0.01	0.9899

Tabla 2: Significancia para el modelo de *Winter Aditivo* sobre el conjunto de datos EJ2.SEMANAL4

Observando la tabla, vemos que tanto para la constante de suavizado para la tendencia como para la estacionalidad no son significativos, es decir, no se rechaza el test α_2 y α_3 igual a 0. Como el estimador α_2 es cercano a 0 y al existir estacionalidad significará que es conveniente diferenciar. Para α_3 será que son índices estacionales deterministas.

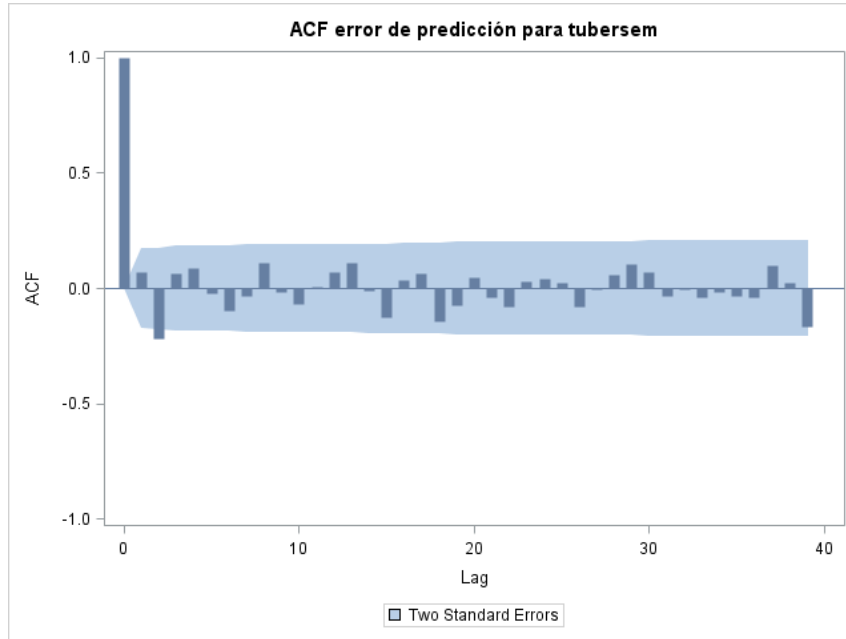


Figura 7: Gráfico de autocorrelaciones (correlograma) para los residuales del modelo de *Winter Aditivo* sobre el conjunto de datos EJ2.SEMANAL4

Siguiendo la línea de lo comentado para el ACF del seasonal, vemos que para el retardo 2 de nuevo vuelve a ser una autocorrelación muy alta que indicará que solo se validará si no encontramos uno mejor. Para las autocorrelaciones para los periodos estacionales no se observa valores altos.

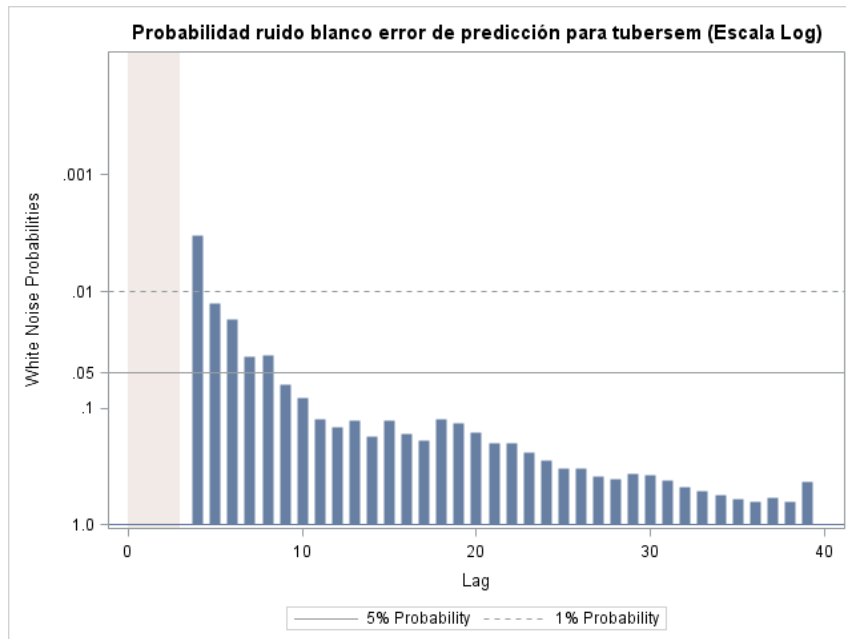


Figura 8: Gráfico sobre el contraste de ruido blanco para los residuales del modelo de *Winter Aditivo* sobre el conjunto de datos EJ2.SEMANAL4

Siguiendo el análisis de este modelo para determinar si es un modelo con ruido blanco o no, vemos en dicha gráfica que de nuevo se rechaza para los primeros retardos y por tanto no será ruido blanco, algo indeseable para validar el modelo.

Por último analizaremos el modelo Winter multiplicativo.

(b.3) Winter Multiplicativo

```
proc esm data = EJ2.SEMANAL4 plot = (errors wn forecasts ACF) print = (all states);
  id DATE interval = WEEK4.6;
  forecast TUBERSEM / model = WINTERS;
run;
```

Código Fuente 4: Ajuste de un modelo de *Winter Multiplicativo* sobre el conjunto de datos EJ2.SEMANAL4

Para comenzar se adjunta la tabla de significancia, la cual se muestra en la Tabla 3.

Método de Winters (Multiplicativo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.57964	0.05438	10.66	<.0001
Peso de tendencia	0.0010000	0.0078088	0.13	0.8983
Peso estacional	0.28227	0.07016	4.02	<.0001

Tabla 3: Significancia para el modelo de *Winter Multiplicativo* sobre el conjunto de datos EJ2.SEMANAL4

Observamos en la tabla anterior, que en este caso si se rechaza para la constante de suavizado estacional con un pvalor < 0,001 y para el nivel. Sin embargo, para la constantes de suavizado para la tendencia, no se rechaza la hipótesis de $\alpha_2 = 0$, por lo que concluimos no será adecuado suavizar dicha componente. Vemos que su estimador es próximo a cero, por lo que será conveniente diferenciar o utilizar modelos ARIMA.

En el gráfico adjunto anteriormente, vemos como los primeros retardos son ligeramente menores a los de los otros modelos, donde en ningún caso sobrepasan las bandas. Los retardos para los periodos no son muy altos, quizás solo la autocorrelación 18 pero no es tan influyente en el ajuste.

Para finalizar, vemos el gráfico WN para contrastar si los residuales del modelo son ruido blanco. En este caso, diferenciándose ligeramente con los modelos anteriores, vemos que para α del 0.01 no se rechaza ninguno y solo 3 para α 0.05.

Por último compararemos SSE de cada modelo a modo de información adicional:

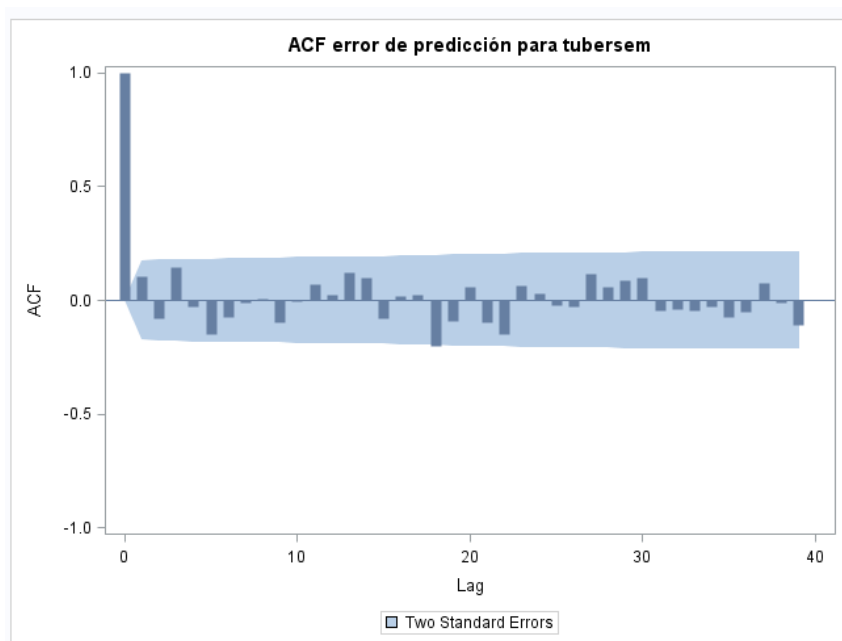


Figura 9: Gráfico de autocorrelaciones (correlograma) para los residuales del modelo de *Winter Multiplicativo* sobre el conjunto de datos EJ2.SEMANAL4

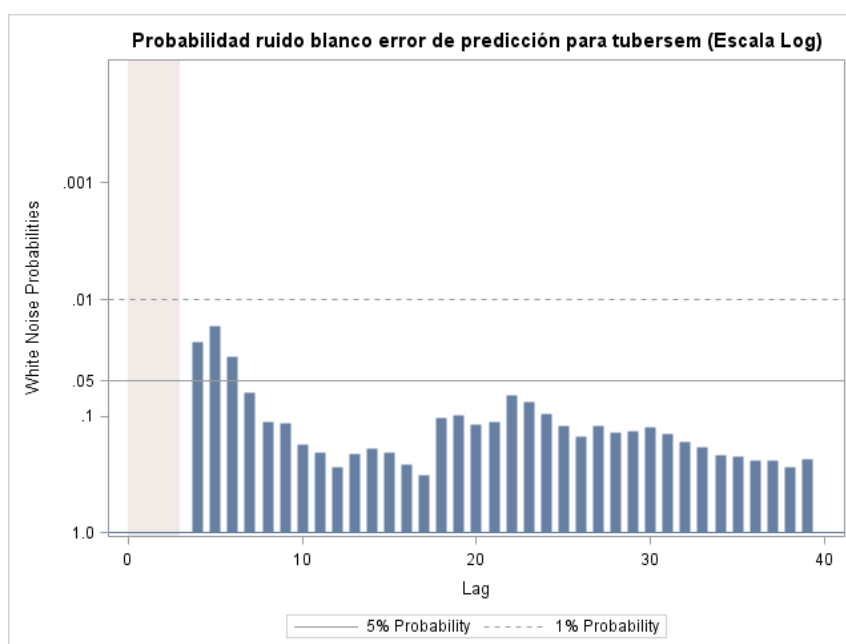


Figura 10: Gráfico sobre el contraste de ruido blanco para los residuales del modelo de *Winter Multiplicativo* sobre el conjunto de datos EJ2.SEMANAL4

Modelo	<i>SSE</i>
Suavizado Exponencial con Estacionalidad	657065.911
Winter Aditivo	661251.008
Winter Multiplicativo	712280.934

Tabla 4: Relación entre los modelos ajustados y su *Suma de Cuadrados del Error*.

- (c) Elegir el modelo que se considere más apropiado entre los tres del Apartado (b) y con ese modelo dar las predicciones para las próximas 6 observaciones. Justificar la elección del modelo.

```
proc esm data = EJ2.SEMANAL4 outfor = PREDICT lead = 6;
  id DATE interval = WEEK4.6;
  forecast TUBERSEM / model = WINTERS;
run;
```

Código Fuente 5: Ajuste y predicción de las 5 observaciones siguientes de un modelo de *Winter Multiplicativo* sobre el conjunto de datos EJ2.SEMANAL4

Tras realizar un exhaustivo análisis en el apartado b, de los 3 modelos de suavizado exponencial, hemos considerado que el mejor modelo es el multiplicativo de Winter. Esta elección la hemos concluido por los motivos mencionados anteriormente :

- Pvalores de la Tabla 3 (Constante de Suavizado Estacional Significativa)
- Gráfica ACF de la Figura 9 (Autocorrelaciones menores que otros modelos)
- Gráfica WN de la Figura 10 (Retardos con menor probabilidad menor de ser ruido blanco)

Por último, obtenemos las predicciones para las seis siguientes observaciones, las cuales se muestran en la Tabla 5.

Predicciones para la variable tubersem					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
131	Fri, 20 Dec 1991	511.5019	74.8900	364.7203	658.2836
132	Fri, 17 Jan 1992	607.9537	91.0373	429.5239	786.3836
133	Fri, 14 Feb 1992	593.2792	99.9137	397.4519	789.1064
134	Fri, 13 Mar 1992	542.8197	104.3162	338.3637	747.2757
135	Fri, 10 Apr 1992	546.4519	113.6206	323.7597	769.1441
136	Fri, 8 May 1992	561.1799	123.9316	318.2784	804.0814

Tabla 5: Predicciones para las próximas 6 observaciones basadas en el modelo de *Winter Multiplicativo*.

- (d) Utilizando en el ajuste solamente los datos hasta el final de 1990 que no incluyan ningún caso de 1991, calcular los errores de predicción para el año 1991 y su correspondiente SSE_p (suma de s errores al cuadrado correspondientes a predicciones $\{1, 2, \dots, s\}$ pasos hacia adelante) para los tres modelos del Apartado (b). Comentar si la elección hecha en el apartado Apartado (c) está de acuerdo con los resultados obtenidos en este caso al comparar la capacidad de predicción de los distintos modelos para el año 1991. Adjuntar el programa con el lenguaje control que hayáis utilizado en este apartado.

Para calcular el SSE_p , primero explicaremos resumidamente qué es y en qué basa. El SSE_p es el cálculo de una medida para comparar la capacidad de predicción de un modelo de series temporales. Se trata de predecir observaciones de una serie estacional de periodo s , para medir la capacidad de predicción de un modelo ajustado a dicha serie.

Si se dispone de n observaciones en total, x_1, x_2, \dots, x_n se reservan las últimas k observaciones, donde k es un múltiplo de s . Para el ajuste se utilizan m observaciones ($m = n - k$) y la medida se obtiene sumando los cuadrados de los residuales $\{1, 2, \dots, k\}$ pasos hacia adelante. Esto se define en la Ecuación 1.

$$\begin{aligned} SSE_p &= \sum_{j=1}^k (x_{m+j} - x_m(j))^2 \\ &= \sum_{j=1}^k e_m(j)^2 \\ &= e_m(1)^2 + e_m(2)^2 + \dots + e_m(k)^2 \end{aligned} \quad (1)$$

En la Tabla 6 se muestran los valores predichos, los errores de predicción, y el SSE_p acumulado para cada una de las 13 observaciones predichas, traduciéndose finalmente en el SSE_p en la observación 130.

Obs	date	tubersem	seasonal	add	mult	errorseasonal	erroradd	errormult	SSEseasonal	SSEadd	SSEmult
118	Fri, 21 Dec 1990	453	538.70	527.94	535.99	-85.699	-74.943	-82.988	7344.37	5616.42	6887.02
119	Fri, 18 Jan 1991	559	683.36	672.69	629.72	-124.361	-113.694	-70.717	22809.94	18542.74	11887.98
120	Fri, 15 Feb 1991	620	666.69	656.13	595.44	-46.692	-36.128	24.557	24990.10	19847.96	12491.03
121	Fri, 15 Mar 1991	551	599.91	589.46	562.34	-48.914	-38.457	-11.344	27382.64	21326.88	12619.71
122	Fri, 12 Apr 1991	577	618.69	608.34	559.36	-41.691	-31.343	17.644	29120.75	22309.29	12931.03
123	Fri, 10 May 1991	667	629.25	619.01	563.97	37.754	47.991	103.033	30546.09	24612.38	23546.79
124	Fri, 7 Jun 1991	625	574.58	564.45	535.86	50.420	60.546	89.135	33088.27	28278.25	31491.89
125	Fri, 5 Jul 1991	534	469.36	459.34	461.26	64.642	74.658	72.743	37266.90	33852.07	36783.39
126	Fri, 2 Aug 1991	483	397.69	387.79	411.09	85.308	95.214	71.906	44544.44	42917.76	41953.88
127	Fri, 30 Aug 1991	525	483.47	473.67	473.22	41.529	51.325	51.783	46269.13	45552.06	44635.38
128	Fri, 27 Sep 1991	487	555.03	545.34	504.62	-68.030	-58.341	-17.618	50897.17	48955.73	44945.77
129	Fri, 25 Oct 1991	443	578.71	569.12	554.93	-135.705	-126.118	-111.932	69313.06	64861.54	57474.49
130	Fri, 22 Nov 1991	495	450.06	440.56	447.37	44.940	54.439	47.632	71332.70	67825.13	59743.29

Tabla 6: Suma de Cuadrados del Error de Predicción acumulada para 13 observaciones hacia delante para el modelo de *Suavizado Exponencial con Estacionalidad*, modelo de *Winter Aditivo* y modelo de *Winter Multiplicativo*.

De manera resumida, en la Tabla 7 se adjunta la capacidad de predicción para los 3 modelos estudiados, donde observamos que el modelo con menor error de predicción para las 13 próximas observaciones es el modelo de *Winter Multiplicativo* cuyo valor $SSE_p = 59743,29$ es el menor de todos ellos..

Obs	date	tubersem	seasonal	add	mult	errorseasonal	erroradd	errormult	SSEseasonal	SSEadd	SSEmult
1	Fri, 22 Nov 1991	495	450.060	440.561	447.368	44.9404	54.4388	47.6319	71332.70	67825.13	59743.29

Tabla 7: Suma de Cuadrados del Error de Predicción $k = 13$ observaciones hacia delante para el modelo de *Suavizado Exponencial con Estacionalidad*, modelo de *Winter Aditivo* y modelo de *Winter Multiplicativo*

El código fuente utilizado para el cálculo de la suma de cuadrados del error de predicción k pasos hacia adelante (SSE_p) se ha incluido en este documento. En concreto el Código Fuente 6 se refiere a la generación del conjunto de datos con el cual ajustar el modelo. Es decir, sin las observaciones que se van a predecir, para después comparar los resultados (esta estrategia de reserva de observaciones se conocer como *Holdout*). El siguiente paso es el ajuste y predicción de las $k = 13$ observaciones siguientes para los modelos. Esto se hace en el Código Fuente 7, Código Fuente 8, y Código

Fuente 9. Finalmente, el cálculo de la *Suma de Cuadrados del Error de Predicción* se lleva a cabo en el Código Fuente 10.

```
data sem4reducido;
  set sem4;
  if _n_ < 118;
run;
```

Código Fuente 6: Generación del conjunto de datos con el cual se ajustar el modelo.

```
proc esm data = sem4reducido outfor = predsea lead = 13;
  id date interval = week4.6;
  forecast tubersem / model = SEASONAL;
run;

data predicseasonal;
  set predsea;
  keep predict;
  rename predict = seasonal;
run;
```

Código Fuente 7: Ajuste del modelo de *Suavizado Exponencial con Estacionalidad* y cálculo de las predicciones $k = 13$ pasos hacia delante.

```
proc esm data = sem4reducido outfor = predadd lead = 13;
  id date interval = week4.6;
  forecast tubersem / model = ADDWINTERS;
run;

data predicadd;
  set predadd;
  keep predict;
  rename predict = add;
run;
```

Código Fuente 8: Ajuste del modelo de *Winter Aditivo* y cálculo de las predicciones $k = 13$ pasos hacia delante.

```
proc esm data = sem4reducido outfor = predmul lead = 13
  plot = (errors wn forecasts ACF) print = (all states);
  id date interval = week4.6;
  forecast tubersem / model = WINTERS;
run;

data predicmult;
  set predmul;
  keep predict;
  rename predict = mult;
run;
```

Código Fuente 9: Ajuste del modelo de *Winter Multiplicativo* y cálculo de las predicciones $k = 13$ pasos hacia delante.

```

data errores;
  merge sem4 predicseasonal predicadd predicmult;
  errorseasonal = tubersem - seasonal;
  erroradd = tubersem - add;
  errormult = tubersem - mult;
run;

data sumaerrores;
  set errores;
  if _N_ >= 118 then SSEseasonal + errorseasonal ** 2;
  if _N_ >= 118 then SSEadd + erroradd ** 2;
  if _N_ >= 118 then SSEmult + errormult ** 2;
run;

data SSEpd;
  set sumaerrores;
  if _n_ = 130;
run;

proc print data = SSEpd;
run;

```

Código Fuente 10: Cálculo del la *Suma de Cuadrados del Error de Predicción* $k = 13$ pasos hacia delante para todos los modelos ajustados.

- (e) Obtener con el `proc forecast` de *SAS* el SSE_p para el modelo de *Winter Multiplicativo* con las mismas constantes de suavizado y los valores iniciales de los parámetros lo más próximos posible a los obtenidos en el Apartado (d) con el `proc esm` para este modelo. Adjuntar el programa con el lenguaje control que hayáis utilizado para obtenerlo.

En la Tabla 8 se muestra la suma de cuadrados del error de predicción obtenida de la misma manera que en el Apartado (d). Esto es fijando los mismos valores iniciales *ASTART* y *BSTART* referidos al término independiente y la pendiente respectivamente. En cuanto a los parámetros estacionales, se ha probado con distintos valores para el atributo *NSSTART* (número de observaciones para estimar los parámetros estacionales) comprobando que el que más se acerca es *NSSTART* = 2, obteniendo un $SSE_p = 62801,35$, muy semejante al obtenido con `proc esm`. El

Obs	predictfor	date	tubersem	error	SSE
1	453.203	Fri, 22 Nov 1991	495	41.7972	62801.35

Tabla 8: Cálculo de la *Suma de Cuadrados del Error de Predicción* (SSE_p) $k = 13$ pasos hacia adelante por el método básico (1) con `proc forecast` para el modelo de *Winter Multiplicativo*

En el Código Fuente 11 se incluyen las sentencias utilizadas para el cálculo de la suma de cuadrados de predicción $k = 13$ pasos hacia adelante con el `proc forecast`.

Paralelamente a la metodología indicada en el Código Fuente 11, se puede seguir otra diferente para llegar a los mismos resultados que elimina la labor de probar diferentes valores para el parámetro *NSSTART*. Esta se incluye en el Código Fuente 12. Los resultados que se obtienen siguiendo estrategia se muestran en la Tabla 9. Como vemos en la última fila, los resultados son equivalentes.

```

proc forecast data = SEM4 OUT = RESULFOR outest = ESTAD weight = (0.55193, 0.001, 0.37214)
  interval = week4.6 LEAD = 13 OUTRESID start = 117 astart = 567.734169 bstart = -0.838957480
  method = WINTERS seasons = 13;
  var TUBERSEM;
run;

data ssepfors;
  set RESULFOR;
  IF _TYPE_ = 'RESIDUAL' AND (_N_ >= 118)
  THEN SUM2 + tubersem * tubersem;
run;

PROC PRINT data = ssepfors;
run;

```

Código Fuente 11: Cálculo de la *Suma de Cuadrados del Error de Predicción* (SSE_p) $k = 13$ pasos hacia adelante por el método básico (1) con `proc forecast` para el modelo de *Winter Multiplicativo*

Obs	_TYPE_	_LEAD_	tubersem	SUM2
126	FORECAST	5	570.004	14439.62
127	RESIDUAL	5	6.996	14488.56
128	FORECAST	6	573.909	14488.56
129	RESIDUAL	6	93.091	23154.41
130	FORECAST	7	533.356	23154.41
131	RESIDUAL	7	91.644	31553.10
132	FORECAST	8	450.001	31553.10
133	RESIDUAL	8	83.999	38609.01
134	FORECAST	9	399.815	38609.01
135	RESIDUAL	9	83.185	45528.78
136	FORECAST	10	454.875	45528.78
137	RESIDUAL	10	70.125	50446.28
138	FORECAST	11	486.295	50446.28
139	RESIDUAL	11	0.705	50446.78
140	FORECAST	12	545.993	50446.78
141	RESIDUAL	12	-102.993	61054.34
142	FORECAST	13	453.203	61054.34
143	RESIDUAL	13	41.797	62801.35

Tabla 9: Cálculo de la *Suma de Cuadrados del Error de Predicción* (SSE_p) $k = 13$ pasos hacia adelante por el método alternativo (2) con `proc forecast` para el modelo de *Winter Multiplicativo*

```
PROC FORECAST DATA = sem4reducido OUT = RESULfor outest = estad weight = (0.55193, 0.001, 0.37214)
  interval = week4.6 LEAD = 13 OUTRESID ASTART = 567.734169 BSTART = -0.838957480 NSSTART = 2
  METHOD=WINTERS SEASONS=13;
  var tubersem;
run;

data predmultfor;
  set resulfor;
  keep tubersem;
  rename tubersem=predictfor;
run;

data erroresfor;
  merge predmultfor sem4;
  error = tubersem - predictfor;
run;

DATA sumaerroresfor; SET erroresfor;
  IF _N_ >= 118 THEN SSE+error**2;
RUN;

data SSEpfor;
  set sumaerroresfor;
  if _n_=130;
run;

PROC PRINT data=SSEpfor;
run;
```

Código Fuente 12: Cálculo del error de predicción SSE_p mediante el `proc forecast` por el método alternativo (2).

- (f) Ajustar un modelo para la serie $\{X_t\}$ con el módulo Time Series Forecasting System de SAS razonando porqué se ha elegido. Utilizar el modelo elegido para predecir valores futuros de esta serie y establecer la comparación con los seis valores obtenidos en el Apartado (c) junto con sus bandas de predicción.

Para decidir que modelo utilizar hemos comparado el multiplicativo con el modelo estacional, realizando el mismo proceso que en el Apartado (c). Los resultados no han sido concluyentes porque habia una similitud muy alta entre ambos. En el modelo estacional eran significativas la constante que suaviza el nivel y la que suaviza la componente estacional, y en el de Winter multiplicativo ocurría igual, y la que suaviza la tendencia no era significativa y por tanto no útil en el modelo. La única mínima diferencia es el *SSE*, que resultaba menor para el multiplicativo, siendo 417937,2 frente a 417943,1 del estacional.

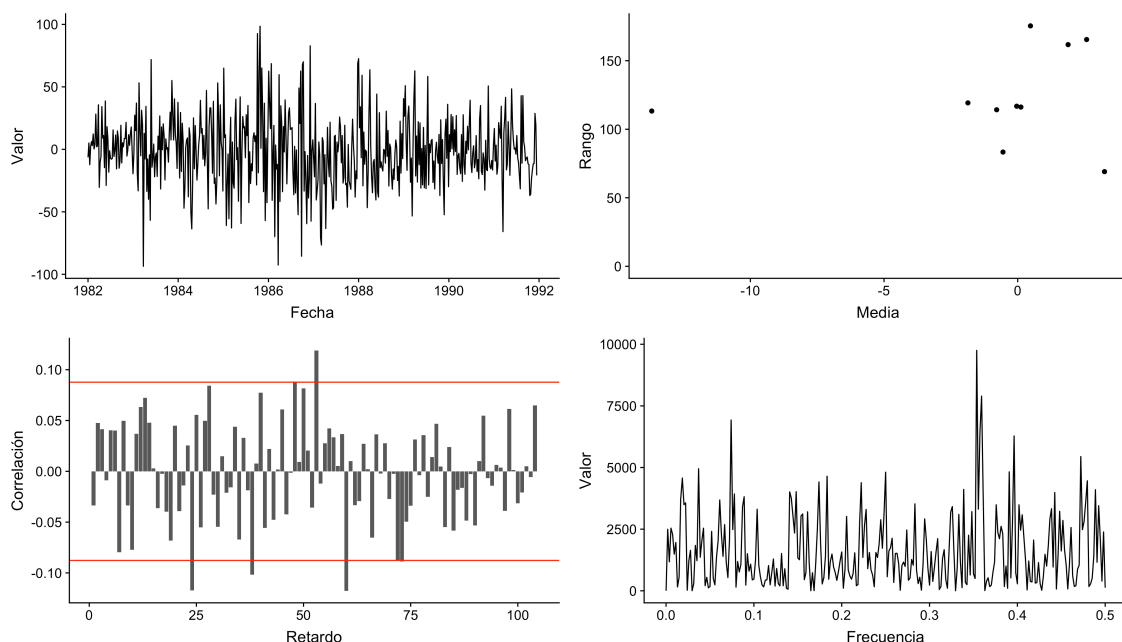


Figura 11: Gráfico de la serie, gráfico de dispersión *rango-media*, *correlograma* y *periodograma* de la serie generada por los residuales del modelo de *Winter Multiplicativo* ajustado al conjunto de datos EJ2.SEMANAL

A continuación se adjunta la tabla de predicciones semanales para el modelo de *Winter Multiplicativo*, obtenida con el siguiente código SAS del Código Fuente 13.

```
/* WORK._F1 es el dataset generado de predicciones para el modelo de winter multiplicativo*/
data NUEVO;
  set WORK._F1;
  if _n_ > 520;
run;

proc print data = NUEVO;
run;
```

Código Fuente 13: Conjunto de datos donde se almacenan los resultados del modelo de *Winter Multiplicativo*

Los datos que se muestran en la 10 son meramente informativos. Lo interesante de los mismos es la suma de estas predicciones agrupadas en grupos de 4 en cuatro. Esto permite para comprarlas con lo obtenido en el Apartado (c) para la serie temporal de datos agrupados cada 4 semanas. Esto se ha llevado a cabo mediante la ejecución del Código Fuente 14.

Obs	DATE	ACTUAL	PREDICT	UPPER	LOWER
1	Friday, December 20, 1991	.	84.5678	140.2938	28.8417
2	Friday, December 27, 1991	.	85.6792	142.3416	29.0169
3	Friday, January 3, 1992	.	116.0866	175.1397	57.0336
4	Friday, January 10, 1992	.	120.4277	180.6584	60.1970
5	Friday, January 17, 1992	.	128.5994	190.4011	66.7977
6	Friday, January 24, 1992	.	131.9833	194.9709	68.9958
7	Friday, January 31, 1992	.	135.9286	200.1999	71.6572
8	Friday, February 7, 1992	.	150.7714	217.8259	83.7168
9	Friday, February 14, 1992	.	138.4468	204.5349	72.3586
10	Friday, February 21, 1992	.	139.4980	206.5370	72.4590
11	Friday, February 28, 1992	.	145.0588	213.7744	76.3432
12	Friday, March 6, 1992	.	130.9209	198.0696	63.7721
13	Friday, March 13, 1992	.	122.9384	189.5459	56.3308
14	Friday, March 20, 1992	.	140.0543	210.5558	69.5527
15	Friday, March 27, 1992	.	136.1069	206.5918	65.6220
16	Friday, April 3, 1992	.	124.4980	193.4448	55.5513
17	Friday, April 10, 1992	.	137.1446	209.4979	64.7913
18	Friday, April 17, 1992	.	127.0161	197.9007	56.1314
19	Friday, April 24, 1992	.	132.1062	204.8915	59.3208
20	Friday, May 1, 1992	.	136.8402	211.4968	62.1836
21	Friday, May 8, 1992	.	142.6926	219.5531	65.8322
22	Friday, May 15, 1992	.	139.2751	215.9519	62.5982
23	Friday, May 22, 1992	.	133.2683	209.0675	57.4690
24	Friday, May 29, 1992	.	145.4126	225.2790	65.5463

Tabla 10: Predicciones del modelo de *Winter Multiplicativo* para la serie temporal referida a observaciones semanales.

A continuación se muestran la tabla que obtuvimos en el Apartado (c) y en la recientemente obtenida sumando las predicciones semanales. Se puede observar como todas las predicciones obtenidas de la suma entran en la banda de predicción de la serie cada 4 semanas. Esto indica un buen ajuste de predicción con este modelo.

Además, realizamos los mismos pasos para el modelo estacional, y obtenemos que la primera predicción de la suma se sale de la banda de predicción (363,4203), lo cual es otro argumento para seleccionar el modelo multiplicativo frente al estacional ya que este ajusta peor las predicciones.

Obs	_NAME_	date	ACTUAL	PREDICT	LOWER	UPPER
1	tubersem	Fri, 20 Dec 1991	.	511.502	364.720	658.284
2	tubersem	Fri, 17 Jan 1992	.	607.954	429.524	786.384
3	tubersem	Fri, 14 Feb 1992	.	593.279	397.452	789.106
4	tubersem	Fri, 13 Mar 1992	.	542.820	338.364	747.276
5	tubersem	Fri, 10 Apr 1992	.	546.452	323.760	769.144
6	tubersem	Fri, 8 May 1992	.	561.180	318.278	804.081

Tabla 11: Predicciones obtenidas por el modelo de *Winter Multiplicativo*, agrupadas en 4 conjuntos de semanas, para poder ser comparadas con los resultados del Apartado (c). Además, estas incluyen sus respectivas bandas de predicción.

Obs	DATE	PREDICT
1	Fri, 20 Dec 1991	406.7613
2	Fri, 17 Jan 1992	547.2827
3	Fri, 14 Feb 1992	553.9245
4	Fri, 13 Mar 1992	523.5976
5	Fri, 10 Apr 1992	533.1070
6	Fri, 8 May 1992	560.6486

Tabla 12: Predicciones agrupadas en Semanas de 4 para comparar con Apartado (c).

```

/*Comparacion*/

/* Individuales Prediccion (Apartado F)*/
proc print data = NUEVO;
run;

/*Suma Prediccion Cada 4 (ApartadoF) */
proc print data = SEM4F;
run;

/* Prediccion C cada 4 para comparar*/
proc print data = CPRED;
run;

```

Código Fuente 14: Comparación de las predicciones obtenidas de manera semanal frente a las obtenidas por grupos de 4 semanas.

(I) Código Fuente Auxiliar

(I.I) Código Fuente SAS

```

data EJ2.SEMANAL;
    input TUBERSEM @@;
    DATE = intnx('week.6', '01jan1982'd, _n_ - 1);
    format DATE weekdate.;
datalines;
106 132 119 147 160 168 166 160 172 190 167 175 197 108 160 148 202 176 195 155 210 149 192 187
136 148 151 152 125 151 111 124 136 99 143 146 118 129 170 118 152 156 167 165 195 145 160 159
166 161 133 104 113 145 177 191 227 189 159 242 224 202 231 234 115 162 191 215 178 186 170 177
136 245 180 185 164 139 160 165 118 154 120 157 105 133 140 166 151 127 175 137 162 157 184 199
201 217 213 184 233 195 148 109 189 193 171 243 218 259 231 195 209 228 232 243 171 187 234 161
156 172 233 179 216 207 180 190 172 177 219 204 175 168 154 148 149 187 163 133 178 202 237 182
183 243 195 192 186 166 250 216 193 225 156 128 134 222 208 215 178 212 241 169 225 258 170 226
202 162 186 255 193 216 169 229 243 160 211 233 232 204 180 230 219 208 149 151 139 144 131 170
180 172 185 176 252 184 275 325 243 326 266 235 274 204 243 135 153 256 267 258 341 264 310 299
236 312 273 284 206 301 187 320 278 248 281 301 316 291 272 292 323 289 270 303 305 226 207 205
182 205 166 165 263 238 294 189 293 323 264 297 257 248 249 233 224 331 197 176 163 208 215 271
234 239 262 260 260 216 177 258 251 199 186 191 193 238 207 222 245 196 197 173 170 189 185 155
193 153 160 162 165 139 150 131 145 176 143 147 113 164 146 145 125 166 169 161 140 120 119 76
153 204 176 202 184 253 192 211 194 214 195 167 181 207 223 225 178 176 186 210 176 150 235 162
156 196 151 165 163 126 141 125 119 155 119 149 121 148 150 136 109 134 129 151 150 164 120 132
96 155 106 102 84 160 159 195 169 158 202 204 190 165 164 126 174 200 211 177 182 159 198 190
209 149 193 152 196 139 156 134 167 192 119 131 128 139 131 111 131 110 138 149 126 136 162 124
168 155 151 126 81 150 147 90 129 109 149 147 180 165 200 164 183 174 190 151 169 163 136 153
164 135 185 166 149 156 167 174 167 171 187 151 150 138 136 108 112 134 98 112 104 148 170 129
109 138 140 116 148 138 117 184 118 133 113 84 77 121 135 120 138 122 128 171 156 155 185 124
76 150 159 166 137 127 160 153 139 156 187 185 165 165 152 143 158 149 123 104 78 149 108 148
128 132 128 137 135 114 129 109 102 111 122 108 116 133 149 97
;
run;

```

Código Fuente 15: Generación del conjunto de datos EJ2.SEMANAL

(I.II) Código Fuente R

```

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, with.first = FALSE) {
  result <- acf(x, lag.max=n, plot=FALSE)$acf[1:n + !with.first]
  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))
}

PlotTimeSeries <- function(df, seasonality, armonics = c(), lags = MAX_LAG){
  p.a <- ggplot(df) +
    aes(x = index, y = values) +
    xlab("Fecha") +
    ylab("Valor") +
    geom_line()
  p.b <- ggplot(RangeMean(df$values, seasonality)) +
    aes(x = mean, y = range) +
    geom_point() +
    xlab("Media") +
    ylab("Rango") +
    expand_limits(y=0)
  p.c <- ggplot(Correlogram(df$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")
  p.d <- ggplot(Periodogram(df$values)) +
    aes(x = freq, y = spec) +
    xlab("Frecuencia") +
    ylab("Valor") +
    geom_line()

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

```

Código Fuente 16: Generación de conjunto de gráficos (gráfico de la serie, gráfico *rango-media*, *correlograma* y *periodograma*) para análisis descriptivo de las series SEMANAL, SEMANAL4 y SEMANALRESIDUOS.