

Análisis de Series Temporales: Tarea 2 - Ejercicio 2

Alejandro Del Hierro Diez

Gabriel Rodríguez Canal

Miguel Martín Mateos

Sergio García Prado

1 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.

[TODO]

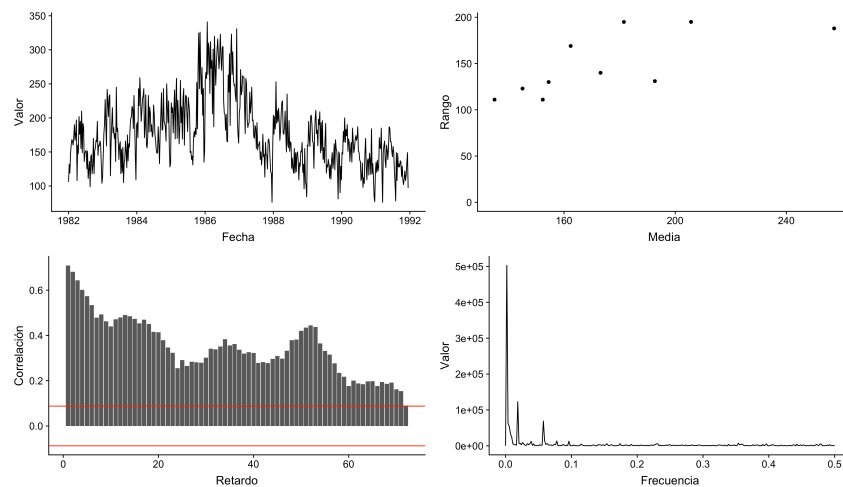


Figura 1: [TODO]

[TODO]

[TODO]

[TODO]

[TODO]

[TODO]

[TODO]

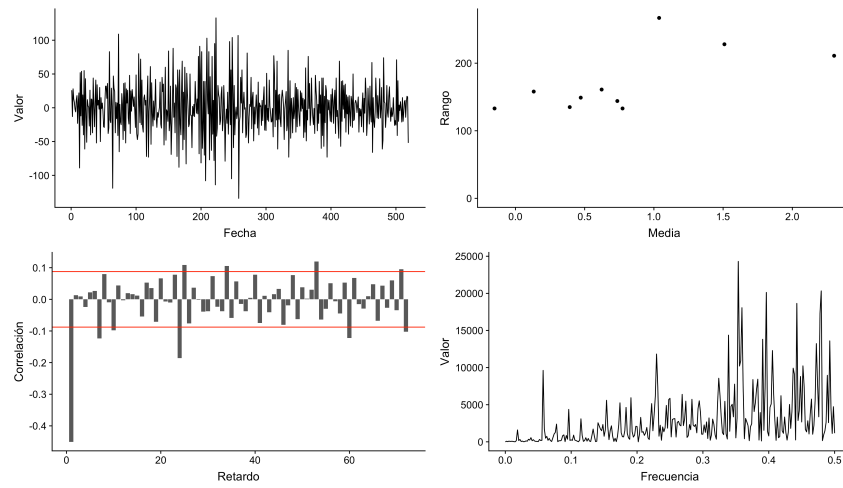


Figura 2: [TODO]

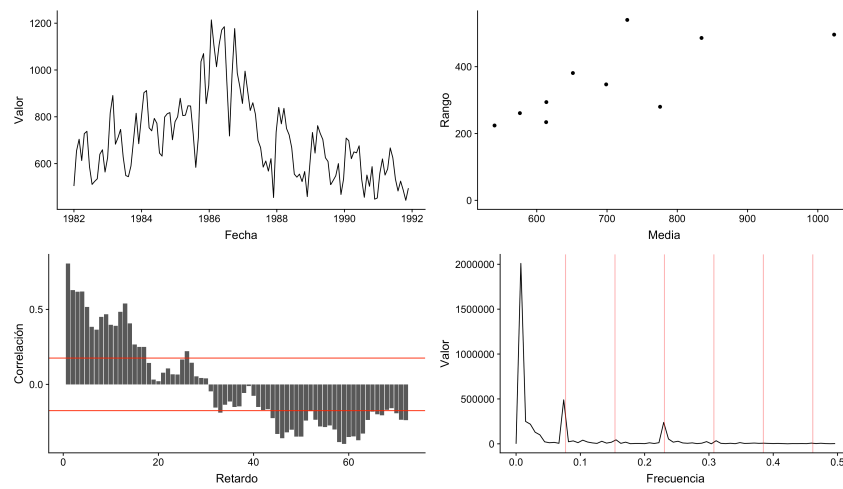


Figura 3: [TODO]

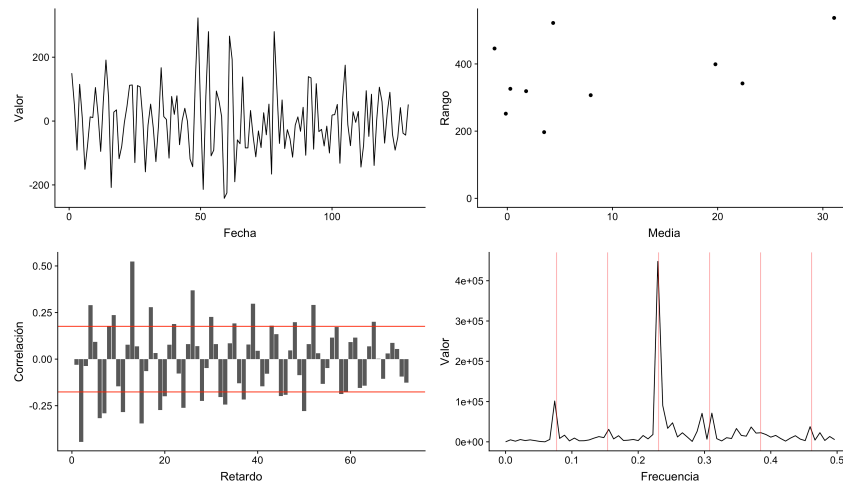


Figura 4: [TODO]

[TODO]

[TODO]

```

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;

```

Figura 5: [TODO]

```

proc expand data=ej2.semanal out=ej2.semanal4 from=week.6 to=week4.6 ;
  id date;
  convert tubersem/ observed=total;
run;

```

Figura 6: [TODO]

```

/* TODO: Describir serie X_t. */

```

Figura 7: [TODO]

```

/* TODO: Describir serie Y_t. */

```

Figura 8: [TODO]

- (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) Seasonal Exponential Smoothing

[TODO]

```

/* TODO: Modelo 1. */
proc esm data=<input-data-set> out=<output-data-set>;
  id <time-ID-variable> interval=<frequency>;
  forecast <time-series-variables>;
run;

```

Figura 9: [TODO]

[TODO]

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

Figura 10: Tabla de significancia. Seasonal

Observamos que para la constante de suavizado estacional no se rechaza su significancia a cualquier nivel ya que el pvalor es 0.99.

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

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.

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) Modelo aditivo de Winter

[TODO]

[TODO]

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.

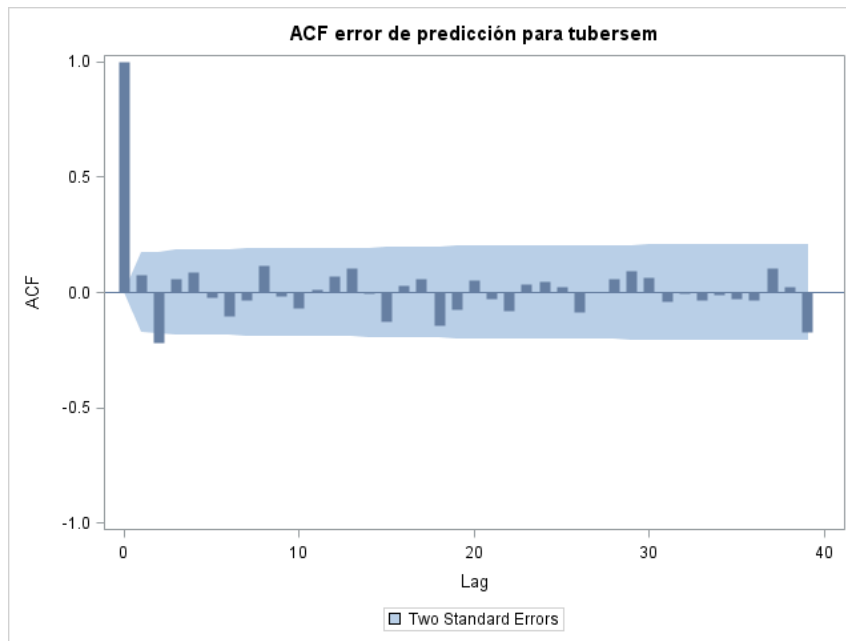


Figura 11: Gráfica ACF Seasonal

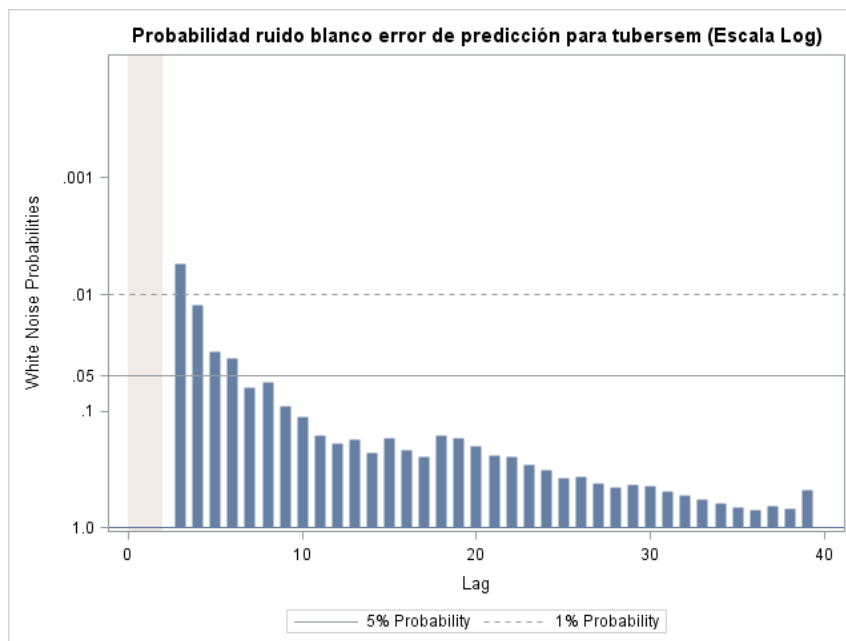


Figura 12: Probabilidad ruido blanco. Seasonal

```

/* TODO: Modelo 2. */
proc esm data=<input-data-set> out=<output-data-set>;
  id <time-ID-variable> interval=<frequency>;
  forecast <time-series-variables>;
run;

```

Figura 13: [TODO]

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.

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

Figura 14: Tabla significancia. Aditivo

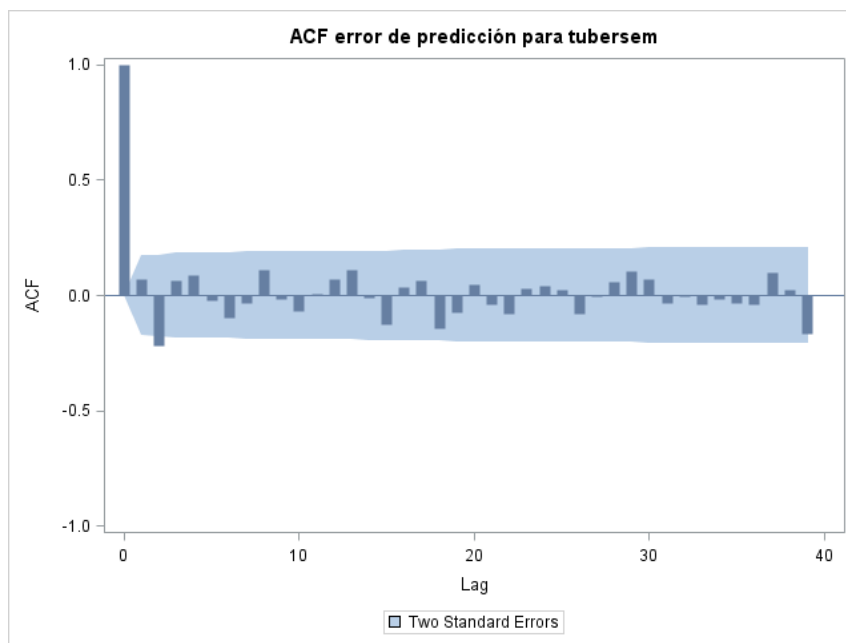


Figura 15: Plot ACF Aditivo. Aditivo

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) Modelo multiplicativo de Winter

[TODO]

[TODO]

Para comenzar se adjunta la tabla de significancia:

Observamos en la tabla anterior, que en este caso si se rechaza para la constante de suavizado estacional con un $p\text{-valor} < 0,001$ 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.

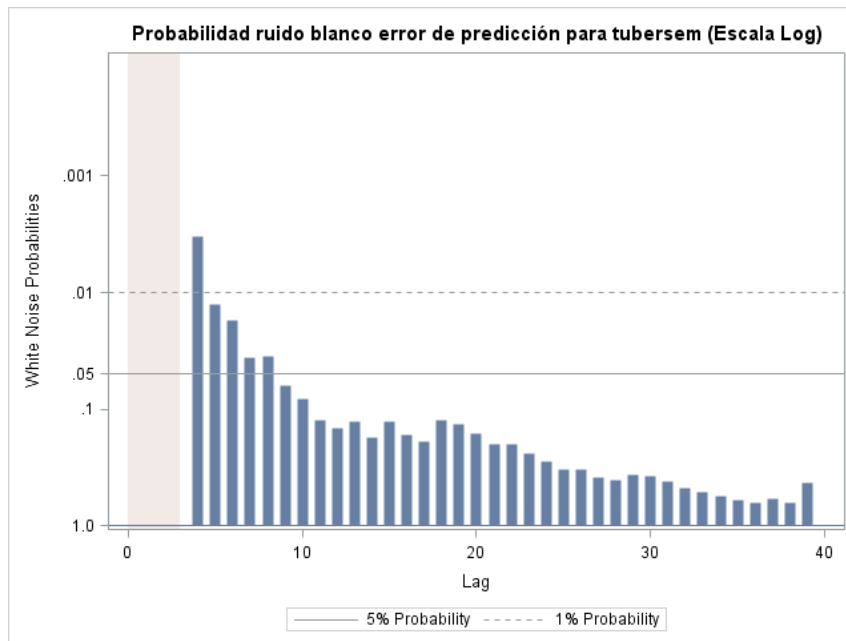


Figura 16: Plot WN Aditivo. Aditivo

```

/* TODO: Modelo 3. */
proc esm data=<input-data-set> out=<output-data-set>;
  id <time-ID-variable> interval=<frequency>;
  forecast <time-series-variables>;
run;

```

Figura 17: [TODO]

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

Figura 18: Tabla Significancia. Multiplicativo

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:

[TODO]

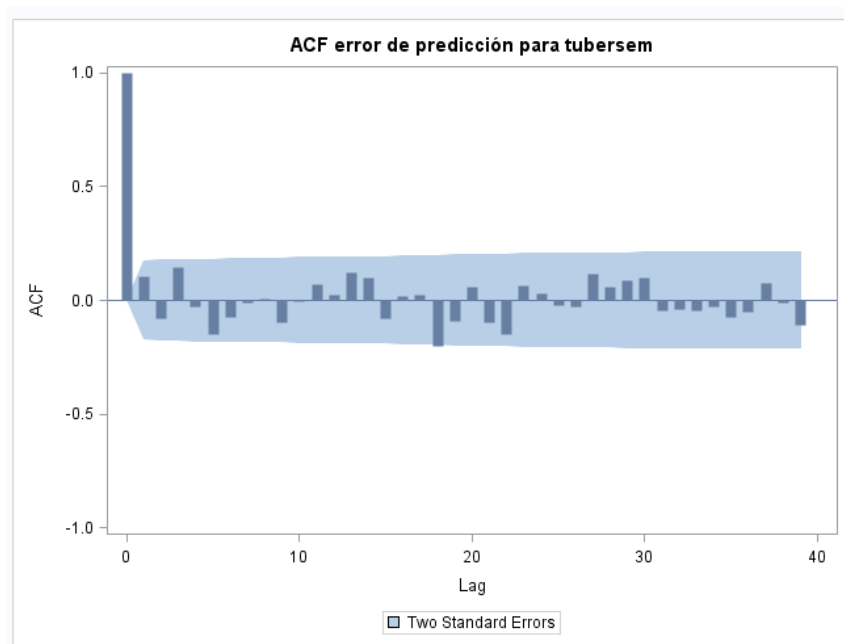


Figura 19: Tabla PLOT ACF. Multiplicativo

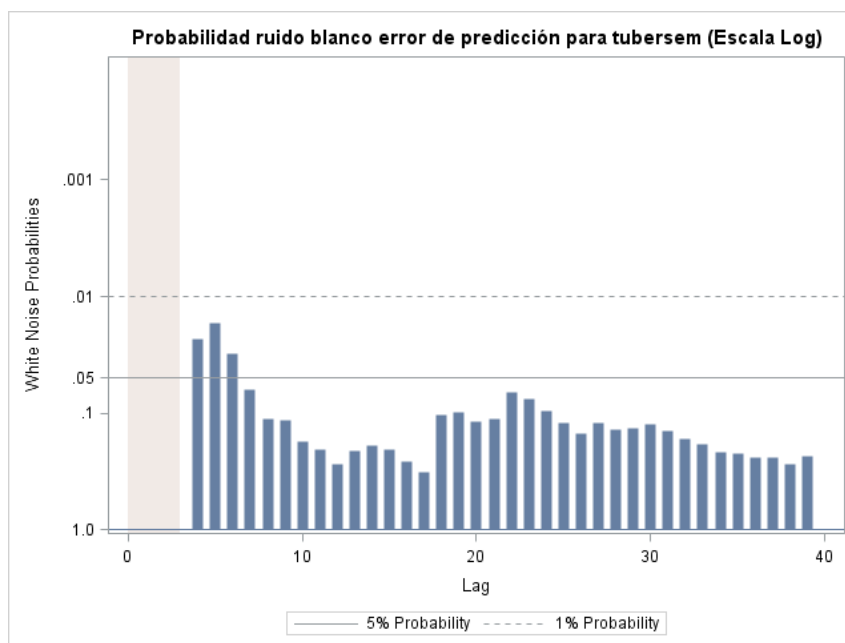


Figura 20: WN.Mul

Error de suma de cuadrados	657065.911
-----------------------------------	-------------------

Figura 21: Tabla SSE.Seasonal

Error de suma de cuadrados	661251.008
-----------------------------------	-------------------

Figura 22: Tabla SSE.ADD

Error de suma de cuadrados	712280.934
-----------------------------------	-------------------

Figura 23: Tabla SSE.MUL

- (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.

[TODO]

```
/* TODO: Realizar predicciones de 6 observaciones siguientes. */
proc esm data=<input-data-set> out=<output-data-set>;
  id <time-ID-variable> interval=<frequency>;
  forecast <time-series-variables>;
run;
```

Figura 24: [TODO]

[TODO]

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 Figura 7(Constante de Suavizado Estacional Significativa)
- Gráfica ACF de la Figura 8 (Autocorrelaciones menores que otro modelos)
- Gráfica WN de la Figura 9(Retardos con menor probabilidad menor de ser ruido blanco)

Por último, obtenemos las predicciones para las seis siguientes observaciones.

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

Figura 25: Tabla Predicciones.Modelo Multiplicativo

[TODO]

- (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 (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. Adjuntad el programa con el lenguaje control que hayáis utilizado en este apartado.

[TODO]

Para calcular el SSE_p , primero explicaremos resumidamente que és y en que 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 un paso, dos, ..., k pasos hacia adelante:

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

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

Figura 26: Tabla Predicciones. Modelo Multiplicativo

A continuación, se adjunta una tabla con los valores predichos, los errores de predicción, y el SSE acumulado para cada observación de las 13 predichas, traduciéndose finalmente en el SSEp 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

Figura 27: Tabla Errores Modelos. Seasonal Add Mul [TODO]

En la tabla siguiente se adjunta la capacidad de predicción para los 3 modelos, donde observamos que el que mejor predirá es el modelo de Winter multiplicativo con un menor SSEp de 59743.29.

[TODO]

[TODO]

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

Figura 28: Tabla Predicciones SSEP. Seasonal Ad Mul

```

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

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

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

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

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

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

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;

```

Figura 29: [TODO]

[TODO]

```
/* TODO: Calcular SSE_p con proc esm para modelo 1. */
```

Figura 30: [TODO]

```
/* TODO: Calcular SSE_p con proc esm para modelo 2. */
```

Figura 31: [TODO]

[TODO]

```
/* TODO: Calcular SSE_p con proc esm para modelo 3. */
```

Figura 32: [TODO]

[TODO]

- (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.

[TODO]

Esta primera tabla se ha obtenido de la misma manera que en el apartado anterior probando a variar el atributo `NSSTART` que finalmente se realizó para 2 obteniendo un SSE_p de 62801.35 que es muy semejante al obtenido con `proc esm`.

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

Figura 15: Tabla Predicciones SSEP Forecast. Multiplicativo

Hay otra manera en la que no hace falta ir modificando el valor en `NSStart`.

[TODO]

[TODO]

[TODO]

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

Figura 33: [TODO]

```

/*
Forma 1
*/
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;

/*
Forma 2
*/

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 SSEpfors;
  set sumaerroresfor;
  if _n_=130;
run;

PROC PRINT data=SSEpfors;
run;

```

Figura 34: [TODO]

```

/* TODO: Calcular SSE_p con proc forecast para modelo 1. */
proc forecast data=past interval=month lead=10 out=pred;
var sales;
id date;
run;

```

Figura 35: [TODO]

- (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.

[TODO]