

Análisis de Series Temporales: Tarea 2 - Ejercicio 1

Alejandro Del Hierro Diez

Gabriel Rodríguez Canal

Miguel Martín Mateos

Sergio García Prado

1 de diciembre de 2018

■ **Archivo:** NE2.txt

■ **Serie:** Número de espectadores de cine por meses desde Enero de 1984 a Diciembre de 2000.

Tal y como se indica al comienzo del documento, en este trabajo se va a tratar de ajustar una serie referida al número de espectadores de cine por meses desde Enero del año 1984 hasta Diciembre del año 2000. Esto se corresponde con un total de 204 observaciones mensuales. La serie completa recoge información de $204/12 = 17$ años acerca del número de espectadores.

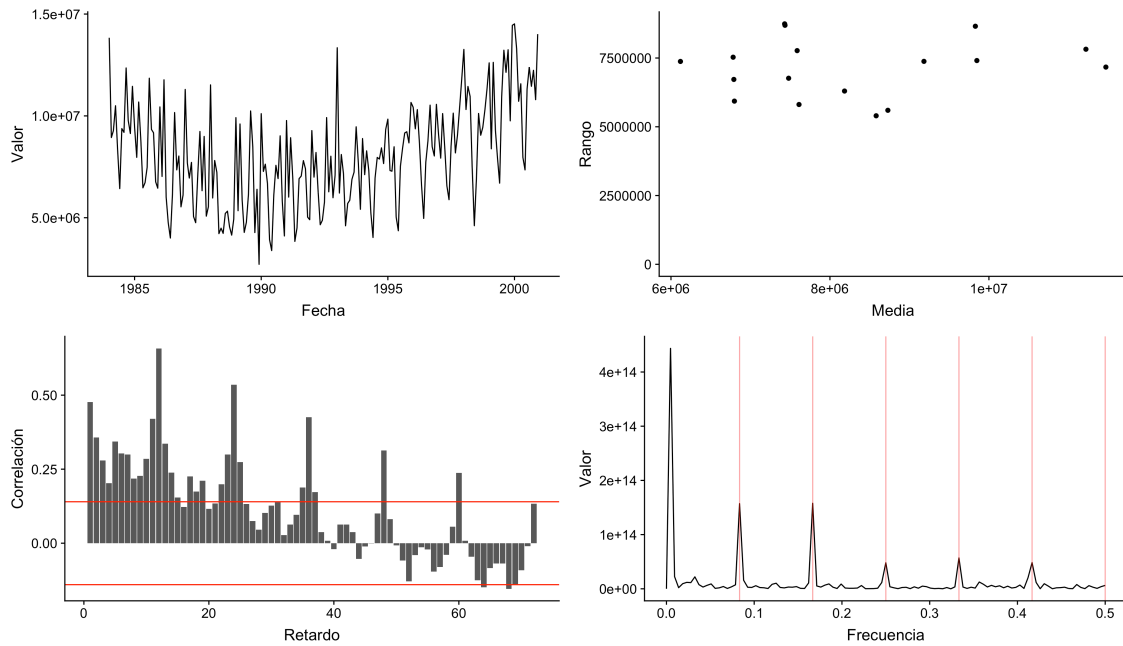


Figura 1: Gráfico de la serie, Diagrama de Dispersión, Correlograma y Periodograma de la serie *Espectadores*.

El primer paso antes de realizar ningún ajuste sobre la serie es conocer la estructura de la misma, lo cual permitirá entender mejor su comportamiento facilitando la realización de una modelización más coherente. Por tanto, en la figura 1 se muestra el gráfico de la serie, el gráfico de dispersión *rango-media*, el correlograma y el periodograma de la misma respectivamente.

Lo más notable en la estructura de la serie es su marcada tendencia parabólica, con su mínimo en torno a los años 1990. Esto nos será de gran ayuda para modelizar la serie, utilizando una modelización cuadrática para la componente de tendencia. También parece verse una componente estacional, que es difícil de estudiar directamente debido a la tendencia. Sin embargo, esta se puede ver bien en el periodograma y correlograma, los cuales comentaremos posteriormente.

En cuanto al diagrama de dispersión *rango-media*, este ha sido generado realizando agrupaciones de longitud igual a la estacionalidad (12 meses). En este gráfico no parece que haya una relación fuerte entre dichas variables. Esto se intuye debido a la pendiente nula que se puede apreciar en el gráfico.

En cuanto al correlograma, en este se aprecia la estructura de correlaciones que indica una componente de tendencia fuertemente marcada. Esta se aprecia en las correlaciones múltiplos de la estacionalidad. Es decir, aquellas correlaciones tales que $i \bmod 12 = 0$. Si nos fijamos únicamente en estas, podemos apreciar un decaimiento marcadamente lineal, lo cual nos confirma la existencia de una componente estacional.

En cuanto al periodograma, se vuelve a confirmar la idea que acabamos de indicar en el anterior párrafo: La serie tiene una marcada estacionalidad de periodo 12. En este caso queda reflejado en la forma determinista en que las frecuencias de la forma $i/12$ con $i \in \{1, 2, \dots, 6\}$ destacan sobre el resto, en especial las dos primeras.

- (a) **Ajustar por mínimos cuadrados un modelo con tendencia más ondas a la serie escribiendo explícitamente las ecuaciones de las ondas incluidas en el modelo. Añadir la descripción completa de la serie residual.**

Una vez descrita la serie, estamos en condiciones suficientes para realizar un ajuste sobre la serie. En este caso la modelización de la tendencia se realizará basándose en un modelo con tendencia (cuadrática) + ondas. Esto lo haremos en dos fases, inicialmente ajustaremos un modelo con todas las ondas posibles, cuyo resultado deberá ser totalmente equivalente al modelo basado en índices que se ajusta en el segundo apartado. Posteriormente, para reducir la complejidad del modelo, eliminaremos aquellas componentes de la onda que no resulten significativos para explicar el modelo.

En cuanto a la componente de la tendencia de la serie, ajustaremos un modelo basado en regresión polinómica de segundo orden, esto es $E[Y_t] = \beta_0 + \beta_1 \cdot t + \beta_2 t^2$ (nótese que no se ha indicado el error aleatorio dado que se ha modelizado respecto del valor esperado). En este modelo utilizaremos como origen la observación 0, por lo tanto, para predecir el valor esperado para la observación $t + 1$ deberemos utilizar el índice $t + 1$.

En cuanto a la modelización de la tendencia, esta se basará en una modelización de ondas como sumas de senos y cosenos con frecuencia angular $\omega_i = \frac{2\pi i}{12}$ con $i \in \{1, \dots, 6\}$. Esto se debe a que la estacionalidad de la serie es de periodo $s = 12$ y por tanto el número de armónicos que podemos utilizar es $s/2$. Por tanto, en este caso $12/2 = 6$. Por lo tanto, la componente regresora basada en ondas será de la forma $\gamma_{i,1} \cos\left(\frac{2\pi i}{12}t\right) + \gamma_{i,2} \sin\left(\frac{2\pi i}{12}t\right)$ para la componente estacional i -ésima y la observación t -ésima. Es importante darse cuenta de que el $\sin(\pi t) = 0$ para todo t entero, por lo que ignoraremos dicho regresor en el modelo.

Entonces, nuestro modelo requiere de la estimación de 3 parámetros referidos a la componente de tendencia y $6 + 6 - 1 = 11$ parámetros referidos a la componente de estacionalidad. Esto nos da un total de $3 + 11 = 14$ parámetros a estimar para tener bien definido el modelo. A continuación se indica el modelo completo:

$$\begin{aligned} E[X_{n+k}] = & \beta_0 + \beta_1 \cdot (n+k) + \beta_2 \cdot (n+k) \\ & + \gamma_{1,1} \cos\left(\frac{2\pi}{12}t\right) + \gamma_{1,2} \sin\left(\frac{2\pi}{12}t\right) \\ & + \gamma_{2,1} \cos\left(\frac{2\pi 2}{12}t\right) + \gamma_{2,2} \sin\left(\frac{2\pi 2}{12}t\right) \\ & + \gamma_{3,1} \cos\left(\frac{2\pi 3}{12}t\right) + \gamma_{3,2} \sin\left(\frac{2\pi 3}{12}t\right) \\ & + \gamma_{4,1} \cos\left(\frac{2\pi 4}{12}t\right) + \gamma_{4,2} \sin\left(\frac{2\pi 4}{12}t\right) \\ & + \gamma_{5,1} \cos\left(\frac{2\pi 5}{12}t\right) + \gamma_{5,2} \sin\left(\frac{2\pi 5}{12}t\right) \\ & + \gamma_{6,1} \cos\left(\frac{2\pi 6}{12}t\right) \end{aligned}$$

Parámetro	Valor	Error	T	pvalor
$\widehat{\beta}_0$	6990872	122379	57.1247	<.0001
$\widehat{\beta}_1$	-28561	3369	-8.4768	<.0001
$\widehat{\beta}_2$	418.68249	29.1143	14.3807	<.0001
$\widehat{\gamma}_{1,1}$	1220484	127717	9.5561	<.0001
$\widehat{\gamma}_{1,2}$	40326	127836	0.3155	0.7528
$\widehat{\gamma}_{2,1}$	-975133	127717	-7.6351	<.0001
$\widehat{\gamma}_{2,2}$	816281	127735	6.3904	<.0001
$\widehat{\gamma}_{3,1}$	647163	127717	5.0672	<.0001
$\widehat{\gamma}_{3,2}$	193953	127717	1.5186	0.1305
$\widehat{\gamma}_{4,1}$	-245261	127717	-1.9204	0.0563
$\widehat{\gamma}_{4,2}$	718853	127711	5.6288	<.0001
$\widehat{\gamma}_{5,1}$	-349795	127717	-2.7388	0.0068
$\widehat{\gamma}_{5,2}$	604979	127708	4.7372	<.0001
$\widehat{\gamma}_{6,1}$	-182690	90306	-2.0230	0.0445
$\widehat{\sigma}^2$	1.66354E12			

Tabla 1: Parámetros ajustados para el modelo basado en ondas con todos los armónicos.

Por lo tanto se ha procedido al ajuste de dicho modelo, el cual se muestran en la tabla 1. Analizando los p-valores (referidos al test acerca de si pueden ser considerados nulos), podemos ver que $\gamma_{1,2}$ y $\gamma_{3,2}$ no rechazan dicha hipótesis. Estos parámetros se refieren a los regresores $\sin\left(\frac{2\pi}{12}t\right)$ y $\sin\left(\frac{2\pi 3}{12}t\right)$ respectivamente. Por lo tanto, vamos a realizar un nuevo ajuste sobre el modelo sin ellos. Para el ajuste de este modelo es necesario estimar 12 parámetros, esto es dos menos que en el modelo anterior. Entonces, el modelo compuesto únicamente con aquellas ondas significativamente distintas de cero se define a continuación:

$$\begin{aligned}
E[X_{n+k}] = & \beta_0 + \beta_1 \cdot (n+k) + \beta_2 \cdot (n+k) \\
& + \gamma_{1,1} \cos\left(\frac{2\pi}{12}t\right) \\
& + \gamma_{2,1} \cos\left(\frac{2\pi 2}{12}t\right) + \gamma_{2,2} \sin\left(\frac{2\pi 2}{12}t\right) \\
& + \gamma_{3,1} \cos\left(\frac{2\pi 3}{12}t\right) \\
& + \gamma_{4,1} \cos\left(\frac{2\pi 4}{12}t\right) + \gamma_{4,2} \sin\left(\frac{2\pi 4}{12}t\right) \\
& + \gamma_{5,1} \cos\left(\frac{2\pi 5}{12}t\right) + \gamma_{5,2} \sin\left(\frac{2\pi 5}{12}t\right) \\
& + \gamma_{6,1} \cos\left(\frac{2\pi 6}{12}t\right)
\end{aligned}$$

En la figura 2 se muestra el gráfico de la serie de predicciones (en color rojo) junto con la serie original (en color gris). Tal y como se puede apreciar, esta se ajusta a los datos. Este ajuste no es perfecto debido a la forma determinista del modelo, que no actualiza las componentes estacionales a lo largo del tiempo. Posteriormente se estudiarán en detalle los residuos para comprobar si la adecuación del modelo que se intuye en la figura 2 es razonable.

En este caso, los parámetros ajustados se muestran en la tabla 2. Tal y como podemos apreciar en este caso todos los pvalores indican que los parámetros son significativamente distintos de cero (tomando un nivel de confianza del 90%). Entonces, procederemos al análisis de la serie de residuales obtenidos para tratar de confirmar el ajuste del modelo. Para ello debemos confirmar que esta serie se refiere a un ruido blanco (no existe correlación entre las distintas observaciones, su media es cero y su varianza constante).

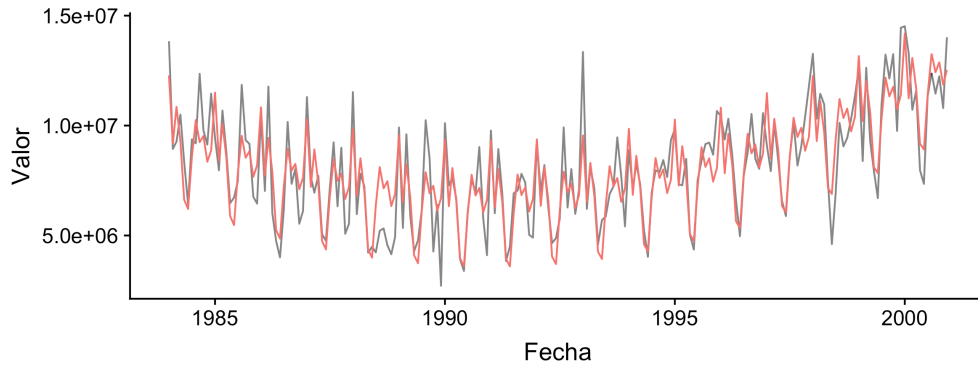


Figura 2: Gráfico de la serie ajustada por el modelo basado en tendencia cuadrática y ondas significativas.

Parámetro	Valor	Error	T	pvalor
$\widehat{\beta}_0$	6993418	122454	57.1107	<.0001
$\widehat{\beta}_1$	-28612	3372	-8.4852	<.0001
$\widehat{\beta}_2$	418.70035	29.1450	14.3661	<.0001
$\widehat{\gamma}_{1,1}$	1220533	127852	9.5464	<.0001
$\widehat{\gamma}_{2,1}$	-975083	127851	-7.6267	<.0001
$\widehat{\gamma}_{2,2}$	816195	127870	6.3830	<.0001
$\widehat{\gamma}_{3,1}$	647212	127851	5.0622	<.0001
$\widehat{\gamma}_{4,1}$	-245212	127851	-1.9179	0.0566
$\widehat{\gamma}_{4,2}$	718825	127845	5.6226	<.0001
$\widehat{\gamma}_{5,1}$	-349746	127851	-2.7356	0.0068
$\widehat{\gamma}_{5,2}$	604965	127843	4.7321	<.0001
$\widehat{\gamma}_{6,1}$	-182666	90401	-2.0206	0.0447
$\widehat{\sigma}^2$	1.66705E12			

Tabla 2: Parámetros ajustados para el modelo basado en ondas con armónicos significativos.

Al igual que se ha indicado anteriormente para el caso de la serie original, en este caso se incluyen un conjunto de diagramas sobre la serie de residuales, la cual se muestra en la figura 3. Al observar el gráfico de la serie vemos que este no tiene tendencia, así como que su media está centrada en el origen. Esto nos hace pensar en ruido blanco, sin embargo debemos fijarnos en el resto de gráficos para confirmar la incorrelación de variables y la varianza constante.

A partir del diagrama de dispersión *rango-media* podemos comprobar que la variación del modelo no está relacionada con la media de este, por lo que podemos confirmar que la varianza se mantiene constante entre toda la serie de residuales. En cuanto a la correlación entre los residuales, tras analizar el correlograma y el periodograma se puede confirmar que no existe ninguna estructura bien definida sobre estos. En el caso del correlograma, la mayoría se encuentran bajo las bandas de confianza, sin representar una estructura claramente marcada.

Si visualizamos el periodograma vemos que no existe ningún conjunto de armónicos que destaque claramente sobre el resto. Gran parte de estos se encuentran distribuidos de manera uniforme a lo largo del soporte de la frecuencia. Esto es otro motivo por el cual asumimos que la serie está bien ajustada y los residuos componen una serie de ruido blanco. A pesar de que existe una leve estructura de correlaciones que refleja estacionalidad, esta se encuentra en todos los casos muy por debajo de las bandas de confianza del 95 %, por lo que ignoraremos dicha estructura por ser demasiado sutil.

Tras el análisis acerca de la serie de residuales del modelo ajustado, podemos asumir que esta se refiere a un ruido blanco, por lo que diremos que el modelo basado en un polinomio cuadrático para modelizar la tendencia y las ondas significativas para la modelización de la estacionalidad de

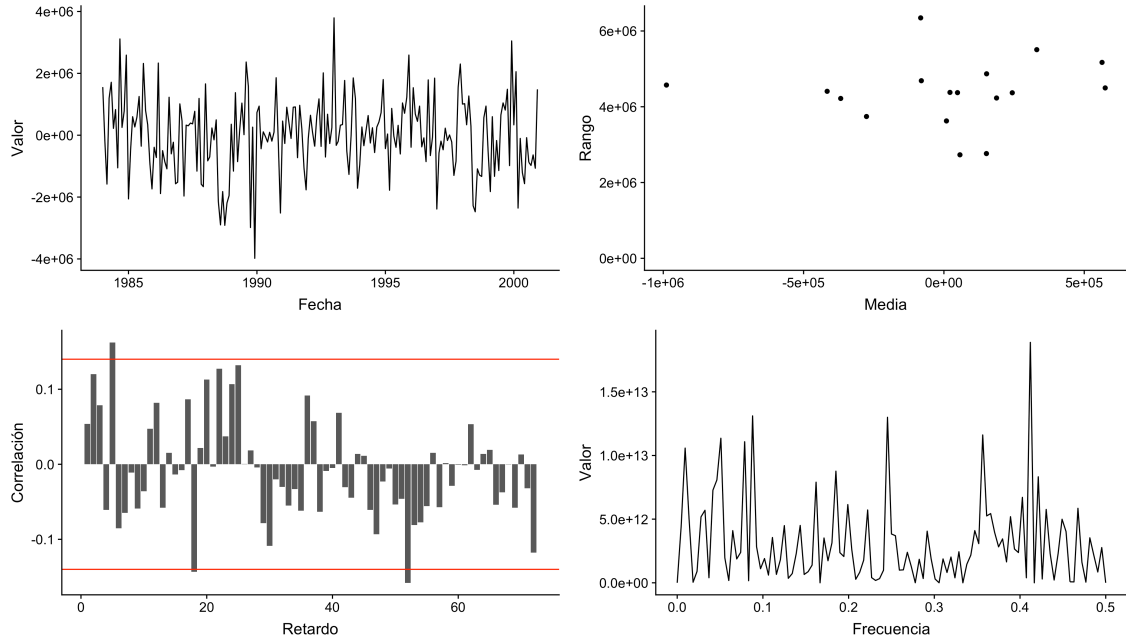


Figura 3: Gráfico de la serie, Diagrama de Dispersión, Correlograma y Periodograma de la serie de residuales tras el ajuste del modelo

periodo 12, ajusta adecuadamente a la serie del número de Espectadores de cine Mensuales entre Enero de 1984 y Diciembre de 2000.

(b) Ajustar por mínimos cuadrados un modelo con tendencia más índices estacionales a la serie y comentar diferencias con el ajuste anterior.

En este apartado se pide ajustar el mismo modelo mediante una modelización basada en *índices estacionales*. Esto es a partir de variables *dummy* tales que toman el valor 1 cuando la observación en cuestión se refiere al periodo estacional de la variable y 0 en caso contrario. A continuación se define más matemáticamente:

$$S_i = \begin{cases} 1 & t \bmod i = 1 \\ 0 & \text{en otro caso} \end{cases}$$

A pesar de que esta modelización es equivalente en cuanto a la cantidad de variabilidad recogida por el modelo y las predicciones que realiza, no lo es desde el punto de vista de los valores de los parámetros (por ser una parametrización diferente). A continuación se muestra el modelo con tendencia cuadrática y representación de la estacionalidad mediante índices:

$$\begin{aligned} E[X_{n+k}] = & \beta_0 + \beta_1 \cdot (n+k) + \beta_2 \cdot (n+k)^2 \\ & + \gamma_1 S_1 + \gamma_2 S_2 + \gamma_3 S_3 + \gamma_4 S_4 \\ & + \gamma_5 S_5 + \gamma_6 S_6 + \gamma_7 S_7 + \gamma_8 S_8 \\ & + \gamma_9 S_9 + \gamma_{10} S_{10} + \gamma_{11} S_{11} \end{aligned}$$

Nótese que se ha eliminado el indicador referido al último índice. Esto se ha hecho dado que la parte del modelo que recoge la tendencia también tiene término independiente, lo cual produce una sobre parametrización y genera problemas de *identificabilidad* entre los regresores. Este término se refiere a que ninguno de estos se pueda escribir como una combinación lineal del resto, lo cual significaría que sobra uno de ellos. Para solucionar esto existen distintas alternativas, sin embargo la herramienta utilizada (**SAS**) opta por eliminar el último regresor del modelo. Entonces, para el ajuste de este modelo necesitamos estimar el valor de 14 parámetros, al igual que el modelo de ondas completas.

Parámetro	Valor	Error	T	pvalor
$\widehat{\beta}_0$	7105638	325747	21.8134	<.0001
$\widehat{\beta}_1$	-28561	3369	-8.4768	<.0001
$\widehat{\beta}_2$	418.68249	29.1143	14.3807	<.0001
$\widehat{\gamma}_1$	2908961	442714	6.5707	<.0001
$\widehat{\gamma}_2$	-303706	442658	-0.6861	0.4935
$\widehat{\gamma}_3$	1249148	442608	2.8222	0.0053
$\widehat{\gamma}_4$	-48819	442563	-0.1103	0.9123
$\widehat{\gamma}_5$	-2469772	442523	-5.5811	<.0001
$\widehat{\gamma}_6$	-3035702	442489	-6.8605	<.0001
$\widehat{\gamma}_7$	-844053	442459	-1.9076	0.0579
$\widehat{\gamma}_8$	1097938	442435	2.4816	0.0139
$\widehat{\gamma}_9$	346445	442416	0.7831	0.4346
$\widehat{\gamma}_{10}$	505551	442403	1.1427	0.2546
$\widehat{\gamma}_{11}$	-783181	442394	-1.7703	0.0783
$\widehat{\sigma}^2$	1.66354E12			

Tabla 3: Parámetros ajustados para el modelo basado en índices.

En la tabla 3 se muestran los parámetros ajustados del modelo. En este caso hay varios que pueden ser considerados distintos de 0 (al igual que ocurría para el caso del modelo de ondas completas). Sin embargo, dado que ya se ha hecho en el anterior modelo, por cuestiones de simplicidad no se va a estudiar el modelo compuesto únicamente por los índices estacionales significativos.

Para comprobar la equivalencia entre el modelo de índices estacionales y el de ondas (incluyendo todos los armónicos), se puede estudiar la varianza del modelo $\widehat{\sigma}^2$, la cual es coincide en ambos (se puede apreciar en las tablas 1 y 3). Esto es porque recogen la misma cantidad de variabilidad.