



**CUNEF**

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS

---

## TÉCNICAS DE PREDICCIÓN: SERIES TEMPORALES

---

Predicción cuota de mercado Colgate y Crest

Aitor Larrinoa Rementería

Noviembre 2021

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Conclusiones</b>	<b>4</b>
<b>3. Datos</b>	<b>5</b>
<b>4. Modelo ARIMA</b>	<b>5</b>
4.1. Análisis de residuos . . . . .	6
4.2. Box-Ljung test . . . . .	6
4.3. Predicciones . . . . .	7
<b>5. Modelo de Intervención</b>	<b>8</b>
5.1. Detección de outliers . . . . .	8
<b>6. Función de transferencia</b>	<b>9</b>

## 1. Introducción

En este informe, se podrá ver la realización de la segunda práctica de la asignatura *Técnicas de predicción: series temporales*.

El trabajo está estructurado en tres grandes partes, la introducción, el cuerpo y las conclusiones. Dentro del cuerpo tenemos 4 secciones. Estas son los datos, el modelo ARIMA, el modelo de intervención y la función de transferencia. Hemos decidido introducir las conclusiones al principio del informe para facilitar la lectura. La parte perteneciente al cuerpo del informe es la más técnica, en la que hablaremos de los resultados obtenidos, daremos breves introducciones a los modelos utilizados, etc. Por otro lado, tendremos las conclusiones que es la sección en la que desmigaremos las conclusiones de los resultados obtenidos durante el cuerpo del informe.

Los capítulos 4, 5 y 6 son los más técnicos del informe. En ellos hemos explicado los pasos a seguir para obtener los resultados que hemos obtenido, además de explicar brevemente los métodos utilizados. Esta parte indica las pinceladas necesarias para poder entender y seguir los métodos utilizados para poder llegar a los resultados con conocimiento de lo que se ha hecho en todo momento.

## 2. Conclusiones

Es importante remarcar la ausencia de datos en el área de marketing. Luego, la dificultad de cuantificar las inversiones en marketing es notable. Aunque si podemos destacar que el incremento en cuota de mercado de Crest, y por tanto disminución de la cuota de mercado de Colgate, se da justo en el momento en el que ADA hace el anuncio a favor de Crest. A partir de entonces, tal y como nos indica la detección de outliers, la cuota de mercado de Crest se mantiene por encima de la de Colgate en todo momento.

Sin embargo, el artículo<sup>1</sup> visible en canvas, resalta que después el anuncio de ADA, aunque Crest disminuye su inversión en marketing, su cuota de mercado sigue aumentando. Y, por otro lado, Colgate aumenta su inversión en marketing pero su cuota de mercado sigue disminuyendo. Esto es un claro ejemplo de que lo que genera estos cambios en la cuota de mercado de ambas empresas es el propio anuncio de ADA.

Atendiendo a los resultados numéricos obtenidos durante la práctica, hemos visto que el modelo ARIMA no nos genera extremadamente buenas predicciones aunque es interesante tener en cuenta los intervalos de confianza ya que no son del todo malas aproximaciones. Por otro lado, hemos analizado y detectado los outliers de ambas series. Esto nos ha permitido detectar el momento del incremento y disminución de cuota de mercado de Crest y Colgate respectivamente en el momento del anuncio de ADA. Finalmente hemos calculado la función de transferencia. Aunque no nos ha proporcionado una información de gran relevancia, hemos observado que la cuota de mercado de Colgate viene explicada por la cuota de mercado de Crest multiplicada por una constante + un ruido adicional.

Terminamos este capítulo de conclusiones, hemos comentado el hecho de que las predicciones del modelo ARIMA no son extremadamente buenas, véase la figura 4. Esto se debe a que entre ambas empresas llegan a abarcar una cuota de mercado de más del 65 % del mercado dentífrico estadounidense. Este hecho no deja mucho margen de maniobra para ambas empresas sea mínimo y por tanto las predicciones son constantes o casi constantes en el caso de Crest.

---

<sup>1</sup>Véase *ASSESSING THE IMPACT OF MARKET DISTURBANCES USING INTERVENTION ANALYSIS* de DEAN W. WICHERN y RICHARD H. JONES

### 3. Datos

Los datos con los que trabajaremos son las cuotas de mercado de las empresas Colgate y Crest durante los años 1958 y 1963. Estos datos se encuentran en el archivo data.xlsx, visible en la plataforma canvas.

Este archivo xlsx tiene cuatro columnas y 276 filas. Las columnas son las siguientes:

- **Year:** Año de cada dato.
- **Week:** Semana del año de cada dato.
- **Colgate:** Cuota de mercado de la empresa Colgate.
- **Crest:** Cuota de mercado de la empresa Crest.

Mostremos ahora una gráfica de ambas series temporales para ver gráficamente los datos.

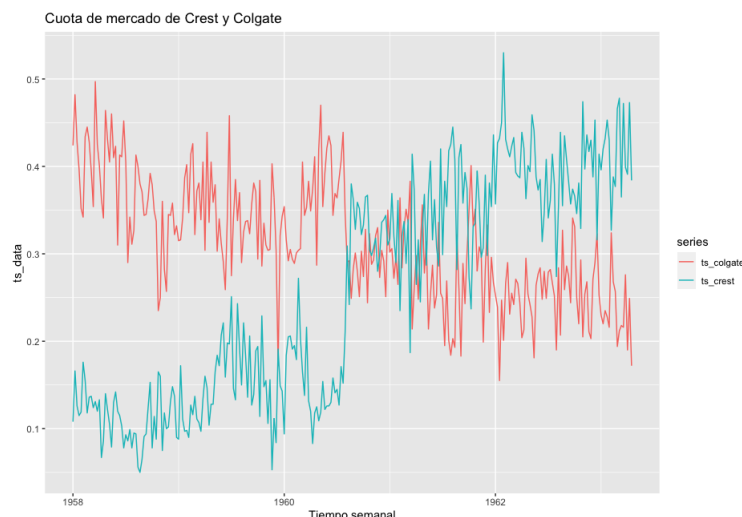


Figura 1: Gráfica de las series temporales

### 4. Modelo ARIMA

Los modelos ARIMA son modelos dinámicos, esto es, son modelos en los que las estimaciones futuras dependen de los datos del pasado. Estos modelos se expresan como  $(p, q, d)x(P, Q, D)$ , donde los parámetros  $p$ ,  $q$  y  $d$  indican las componentes autorregresiva, integrada y media móvil, respectivamente. El primero de los paréntesis indica si la variable que estamos analizando depende de un periodo anterior y el segundo de los paréntesis atiende a la componente estacional del modelo.

Luego, una serie temporal puede venir explicada por la combinación de modelos autorregresivos y de medias móviles, en ese caso tendríamos un modelo ARMA. Además, si la serie es no estacionaria, entonces el modelo que la representa es un modelo ARIMA.

Empezamos a analizar los modelos ARIMA para la series de Colgate y Crest. Para el caso de Colgate, el modelo ARIMA obtenido es  $ARIMA(0,1,1)$ . Esto quiere decir que en el caso de Colgate no tenemos componente estacional.

Por otro lado, el modelo ARIMA obtenido en el caso de Crest es  $ARIMA(3,1,0)(1,0,0)$ . A diferencia del modelo ARIMA con Colgate, en el modelo ARIMA de Crest sí tenemos componente estacional.

#### 4.1. Análisis de residuos

Los análisis de los residuos de ambos modelos se muestran a continuación:

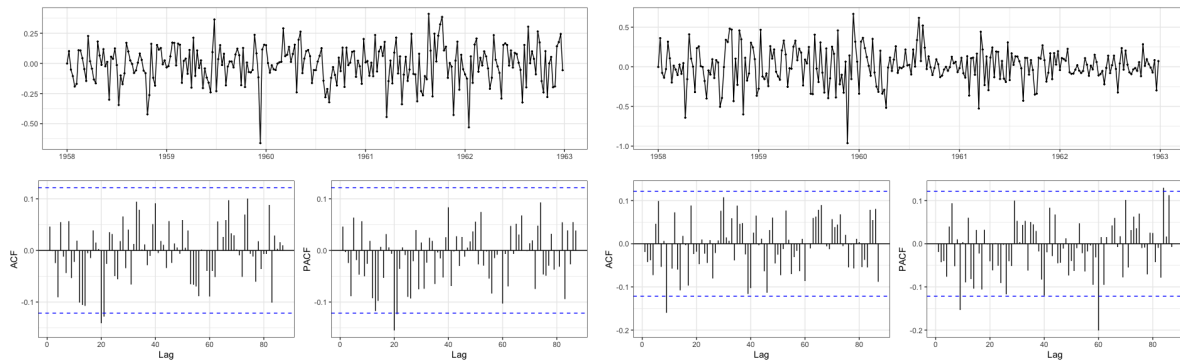


Figura 2: Residuos de los modelos ARIMA

A la izquierda se muestran los residuos del modelo para Colgate y, a la izquierda se muestran los residuos del modelo para Crest. Como se puede observar, los residuos no so ruido blanco. Vemos que existe dependencia lineal entre ellos.

#### 4.2. Box-Ljung test

Por otro lado realizaremos también el Box-Ljung test. Este test estadístico determina si un grupo de autocorrelaciones de una serie temporal son diferentes de 0 o no. Plantendo así el contraste de hipótesis, se tiene que la hipótesis nula es que *los datos se distribuyen de forma independiente* y, la hipótesis alternativa es que *los datos no se distribuyen de forma independiente*.

Realizando dicho contratse, los p-valores que se obtienen son los siguientes:

	Modelo Colgate	Modelo Crest
<b>p-valor</b>	0,7175	0,1597

Cuadro 1: Tabla de p-valores del contraste Box-Ljung para ambos modelos ARIMA

Vistos los resultados, concluimos que en ninguno de los casos se rechaza la hipótesis nula, es decir, en ambos casos se tiene que los residuos no se distribuyen de forma independiente.

### 4.3. Predicciones

Finalmente, para terminar con el modelo ARIMA, veamos las predicciones que tenemos para las últimas 16 semanas de los datos. Haremos uso del dataset de training para realizar la predicción. Después, compararemos esta predicción con los valores de testing. De esta forma veremos cuan buenas son las predicciones obtenidas.

A continuación mostramos ambas predicciones. Únicamente nos hemos centrado en las últimas 16 semanas del dataset ya que ese es nuestro objetivo.

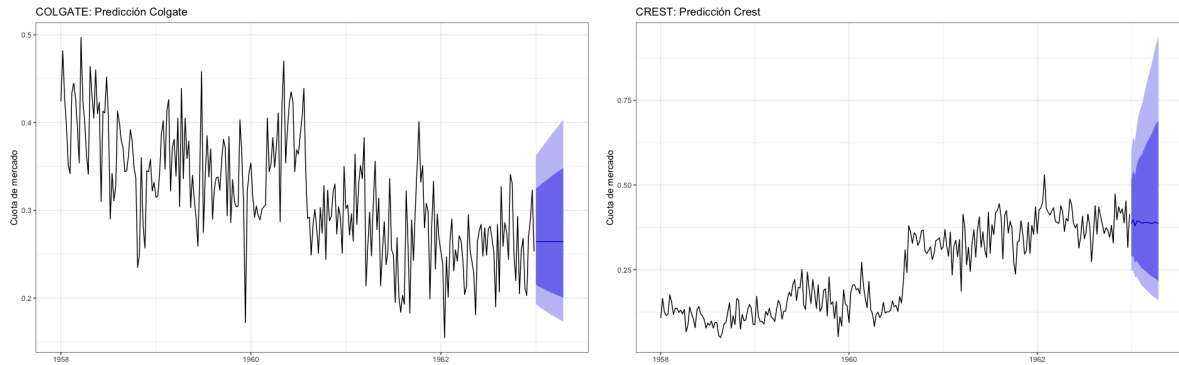


Figura 3: Predicciones de los modelos ARIMA

La región coloreada de color morado oscuro indica el intervalo de confianza del 80 % mientras que la región coloreada de morado claro indica el intervalo de confianza del 95 %.

En las predicciones se puede ver que los modelos no son muy arriesgados .....

Comparando estos resultados con los reales, vemos lo siguiente:

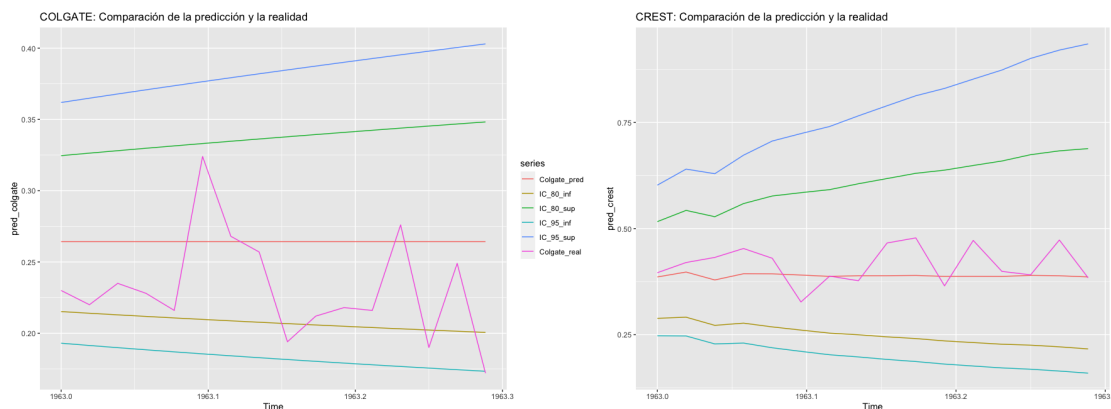


Figura 4: Comparación de las predicciones de los modelos ARIMA

En el caso de Colgate la realidad no se aleja más allá del intervalo de confianza del 95 % pero si supera el intervalo del 80 %, mientras que en el caso de Crest, la realidad no supera el intervalo de confianza del 80 % marcado. Así, podemos decir también que las predicciones para Crest son más precisas que las predicciones de Colgate.

## 5. Modelo de Intervención

Cuando deseamos medir el impacto de un evento determinado en el comportamiento de una serie temporal, deberemos hacer uso de modelos de intervención. Para hallar este tipo de modelos, necesitaremos en primer lugar atender a los outliers de cada una de las series temporales. Después deberemos introducirlos en el modelo ARIMAX correspondiente y el resultado será el modelo de intervención buscado.

### 5.1. Detección de outliers

En primer lugar deberemos detectar los outliers o valores atípicos que tengan nuestras series temporales. Las analizaremos por separado.

En el caso de Colgate, tenemos dos outliers. Los outliers son del tipo “AO” (Aditive outlier) y del tipo “LS” (Level shift). Estos dos outliers han sido notificados en los puntos 102 y 136 de la serie temporal de la cuota de mercado de Colgate, es decir a finales del año 1959 y a mediados del año 1960 (momento del anuncio de ADA). Por otro lado, en el caso de Crest, tenemos tres outliers. Estos son del tipo “AO”, “LS” y del tipo “TC” (Temporary change). Los puntos en los que se han dado son 136, 167 y 197, es decir, a mediados del año 1960 (momento del anuncio de ADA), principios del año 1961 y finales del mismo año respectivamente. El outlier del punto 136 corresponde a un “LS”, el del punto 167 a un “AO” y el del punto 197 a un “TC”. El efecto producido por los outliers los podemos representar en los siguientes gráficos:

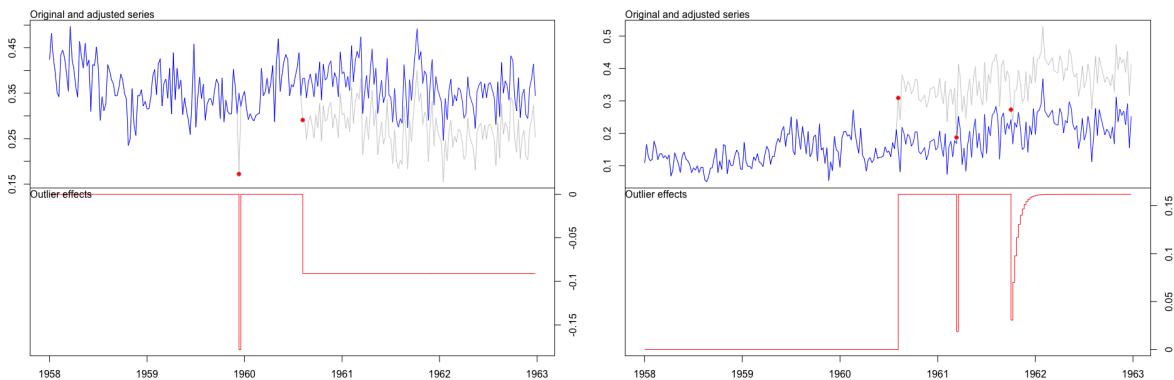


Figura 5: Efecto de los outliers para Colgate y Crest

El gráfico de la izquierda se corresponde con los outliers de Colgate y, el de la derecha con los de Crest. Comparando ambos gráficos, podemos observar que los outliers de Crest muestran un cambio positivo en la cuota de mercado de la empresa mientras que los outliers en Colgate muestran un cambio en la cuota de mercado negativo. El primero de los outliers de Colgate es puntual, pero el segundo es un descenso de la cuota de mercado que no logra recuperarse a lo largo del tiempo. En el caso de Crest tenemos un cambio positivo que después se mantiene en el tiempo, un cambio puntual y un cambio temporal que logra recuperarse a lo largo del tiempo.



## 6. Función de transferencia

Sean dos series  $x_t$  e  $y_t$  tales que:

- Los valores  $x_t$  influyen en los valores  $y_{t+k}$ ,  $k > 0$ .
- La relación entre las variables permanece constante a lo largo del tiempo.
- Se puede escribir  $y_t = y_t^* + n_t$  donde  $n_t$  corresponde a la parte de  $y_t$  no explicada por  $x_t$  linealmente e  $y_t^*$  es la parte explicada por  $x_t$  tal que:

$$y_t^* = \sum_{i=0} \nu_i \cdot x_{t-i} \quad (1)$$

Los valores  $\nu_i$  son los denominados *función de transferencia*. La parte  $n_t$  sigue un ARIMA de parámetros  $(p, q, d)$  y la parametrización de una función de transferencia viene dada por el cociente de dos polinomios.

En nuestro caso, consideraremos que nuestra  $y_t$  corresponderá a la serie temporal de Colgate y, la serie temporal  $x_t$  corresponderá a Crest. Lo primero que debemos hacer será estimar nuestro cociente de polinomios, al que denotaremos por  $\nu(B)$ .

Con el fin de poder encontrar la función de transferencia, es decir, los coeficientes  $\nu_i$ , lo primero que haremos será un modelo ARIMAX. Para ello necesitaremos el orden del modelo ARIMA de la serie temporal de Colgate. Dado que este orden ya ha sido calculado en ?? y es (0,1,1).

Una vez se tiene el modelo ARIMAX realizado, debemos atender a los residuos. Nuestro objetivo será obtener que los residuos sean ruido blanco. Si no lo conseguimos esta vez, deberemos realizar el modelo ARIMAX para las diferencias y comprobar los residuos, así sucesivamente hasta obtener ruido blanco en los residuos. Los residuos obtenidos para el modelo ARIMAX que hemos calculado son:

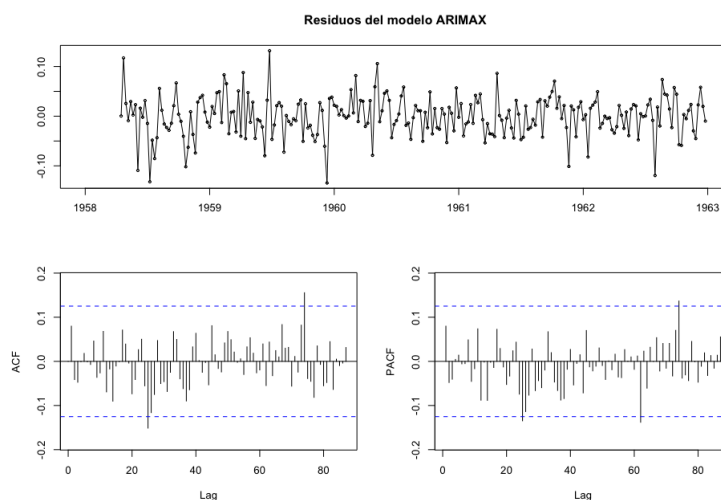


Figura 6: Residuos del modelo ARIMAX

Vemos que hay dos o tres residuos que se escapan muy poco de las líneas azules. Dado que esto es insignificante, consideraremos que los residuos del modelo ARIMAX son 0. Dada esta

condición, estamos en condiciones de analizar los coeficientes obtenidos y hallar los valores de  $b, r, s$  para poder determinar el polinomio necesario para hallar los coeficientes de la función de transferencia. El gráfico de los coeficientes es el siguiente:

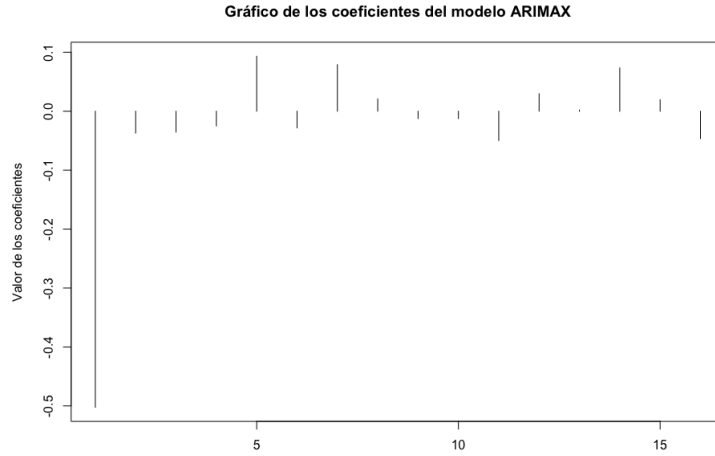


Figura 7: Gráfico de los coeficientes del modelo ARIMAX

Vemos que no hay ningún tipo de decaimiento, luego el valor de  $r$ ,  $b$  y  $s$  es 0. Esto es, el polinomio  $\nu(B)x$  resultante es  $\omega_0 x_t$ . Luego, dicho de otra forma, tenemos que la cuota de mercado de Colgate viene descrita por la siguiente expresión:

$$y_t = \nu_0 x_t + n_t$$

donde  $n_t$  sigue un ARIMA de orden (0,1,1). Como nuestro coeficiente  $\nu_0 = -0,4946$ , nuestra ecuación queda de la siguiente forma:

$$y_t = -0,4946 \cdot x_t + n_t \quad (2)$$

En definitiva, podremos asegurar que la cuota de mercado de Colgate en un instante  $t$ , viene dada por la cuota de mercado de Crest en ese mismo instante ponderada por un valor  $-0,4946$  + una componente de ruido  $n_t$ .