

Práctica final PREDICCIÓN

JORGE CASAN VÁZQUEZ

17 de enero de 2019

EXECUTIVE SUMMARY

El objetivo de este informe final es la viabilidad por parte de nuestra empresa Crest de lanzar una campaña publicitaria que aumente significativamente sus ventas y su cuota de mercado y se afianze, no solo en el mercado estadounidense sino también a nivel europeo. Nuestra tarea se va a basar fundamentalmente en estudiar tal viabilidad y realizar predicciones a través de modelos ARIMA y funciones de transferencia, realizando previamente un análisis de los outliers que respalden tal situación.

En la última semana de 1962 la cuota de mercado de Crest era de 0.414 y de Colgate de 0.253. En base a los datos estadísticos fundamentados en el siguiente análisis, las acciones estratégicas que se deben de implementar para aumentar las ventas en las semanas de 1963 y siguientes son las siguientes:

1. Lanzar un nuevo producto, a nivel local, estudiando la viabilidad de expandir su nicho de mercado más allá de los artículos de aseo personal como un artículo de limpieza, como el detergente. Sabiendo que la sede social de Crest está en Cincinnati (Ohio) podríamos estudiar tal opción.
2. Ofertas y descuentos comerciales por comprar un pack de pasta de dientes Crest e hilo dental de más de un 20%.
3. Acuerdos comerciales con proveedores que oferten tales productos y ofertas en grandes establecimientos comerciales y supermercados.
4. Pocas empresas son conocidas a nivel internacional a principios de los 60. Darse a conocer a través de medios de comunicación en prensa y en radio son fundamentales y establecer un buen sistema logístico de reparto por todo el mercado de EEUU y Europa es fundamental para el abaratamiento de costes de transporte.

Tomado todo ello en su conjunto se mejorará sustancialmente el retorno de la inversión o ROI, el gasto publicitario tendrá un impacto automático en la cuenta de pérdidas y ganancias en el primer trimestre de 1963, según nuestras previsiones ARIMA y ARIMAX.

ANÁLISIS EXPLORATORIO DE DATOS

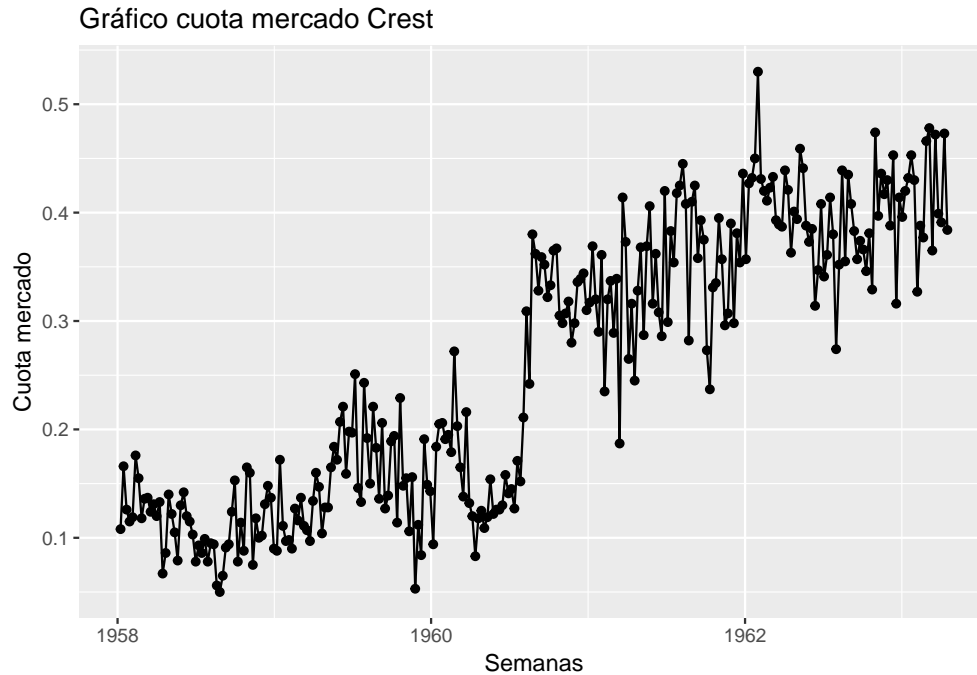
Del conjunto de datos tenemos 276 observaciones y 4 variables, de las cuales 2 de ellas son el año y la semana. Como primer paso, tenemos que homogeneizar su formato para poder comparar entre las dos empresas. Para ello generaremos la fecha desde la primera semana del año 1958 hasta la última del año 1962. Hemos dejado fuera del espacio muestral los datos para el año 1963 puesto que lo que se trata en esta práctica es predecir las 16 semanas del año en cuestión, para las dos cuotas de mercado.

Representando gráficamente ambas cuotas observamos diferencias:

- Crest:
 - Visualmente observamos no estacionalidad, la serie va paulatinamente aumentando. Digno de mención es el salto significativo en un momento del año 1960 que justo parece ocurrir el 1 de Agosto de 1960, en donde en tan solo una semana pasa de tener una cuota de mercado del 0.211 al 0.309, cuando el Consejo de Terapéutica Dental de la “American Dental Association (ADA)” aprobó a Crest como una “ayuda importante en cualquier programa de higiene dental”, el cual lo analizaremos en profundidad con los modelos de intervención.

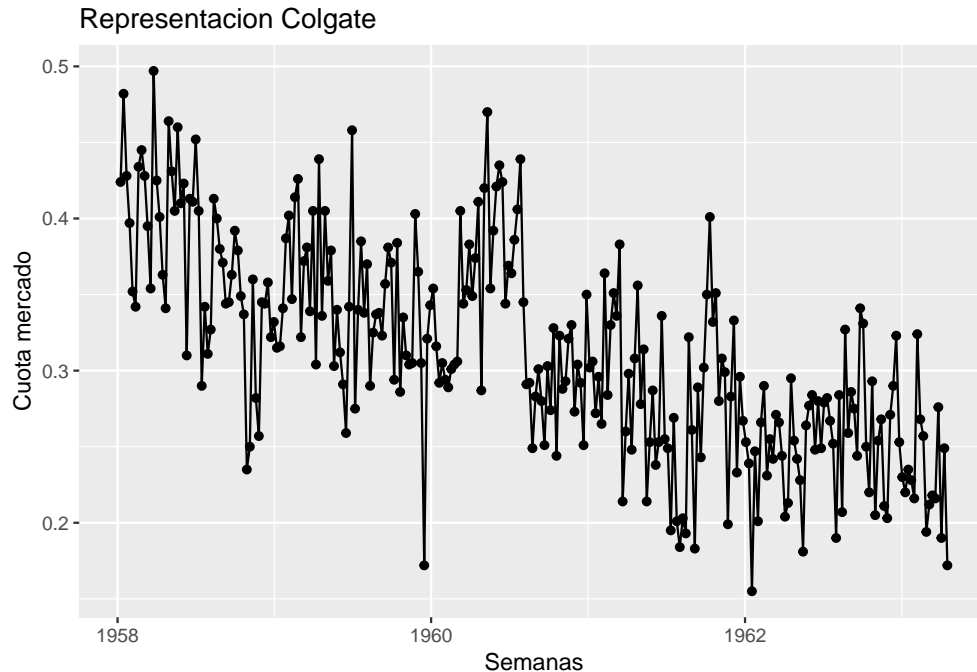
- Nos encontramos ante un escalon, y no un impulso, puesto que la tendencia de la serie temporal no se recupera en cuanto a su media, sino que el gráfico se comporta con una tendencia ascendente, de forma suavizada. Por eso, el orden polinomial que se elige es el wBSt.

```
autoplot(zCuotaCrest) + geom_point() +  
  ylab("Cuota mercado") + ggtitle("Cuota semanal Crest") + xlab("Semanas") +  
  ggtitle('Gráfico cuota mercado Crest')
```



- Colgate:
 - Visualmente observamos no estacionalidad, al igual que con Crest, la serie va paulatinamente descendiendo. Prácticamente en la misma fecha, en donde la ADA mencionaba a Crest como empresa comprometida con la higiene bucal, Colgate pasa de tener en tal sólo una semana una reducción de su cuota de mercado del 0.345 al 0.291.
 - Al igual que con el gráfico de Crest, nos encontramos ante un escalon, y no un impulso, puesto que la tendencia de la serie temporal no se recupera en cuanto a su media, sino que el gráfico se comporta con una tendencia ascendente, de forma suavizada. Por eso, el orden polinomial que se elige es el wBSt.

```
autoplot(zCuotaColgate) + geom_point() +  
  ylab("Cuota mercado") + ggtitle("Gráfico Cuota mercado Colgate") + xlab("Semanas") +  
  ggtitle('Representacion Colgate')
```



· MODELO ARIMA

El modelo ARIMA (modelo autoregresivo integrado de medias móviles) es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

Una serie no estacionaria puede convertirse en estacionaria gracias al modelo ARIMA, en donde a través de la conversión logarítmica podemos convertir la varianza en estacionaria y por otra parte a través de su conversión diferencial podemos convertir la media en estacionaria.

Vamos a determinar el auto arima que deseamos y para ello tendremos que elegir, de entre diferentes modelos, aquel que sea más preciso y para ello escogeremos aquel que tenga menos AIC.

Por una parte, en Crest el mejor es un ARIMA (0,1,1) con un AIC de -864,15, en donde se comprueba que el modelo es estacionario pero no tiene un componente estacional.

```
fit1 = auto.arima(oVentasCrest)
summary(fit1)
```

```
## Series: oVentasCrest
## ARIMA(0,1,1)
##
## Coefficients:
##      ma1
##      -0.6494
## s.e.    0.0448
##
## sigma^2 estimated as 0.002054: log likelihood=434.08
## AIC=-864.15   AICc=-864.1   BIC=-857.04
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE
## Training set 0.003018071 0.04514444 0.03445351 -2.924537 17.27068
```

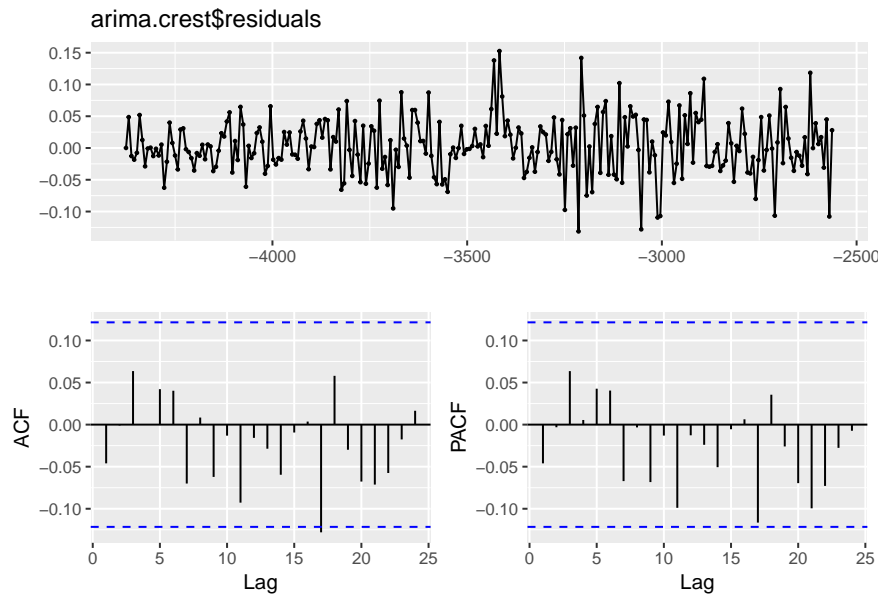
```
##                               MASE          ACF1
## Training set 0.1405405 -0.04605961
```

Por otra parte, en Colgate el mejor ARIMA, también es un ARIMA (0,1,1) con un AIC de -833,14. Al igual que en el anterior, es un modelo estacionario en media y varianza pero sin embargo no tiene componente estacional.

A continuación realizamos el análisis de los residuos con la función de autocorrelación simple (ACF) y con la función de autocorrelación parcial (PACF), viendo que no se salgan de las franjas azules.

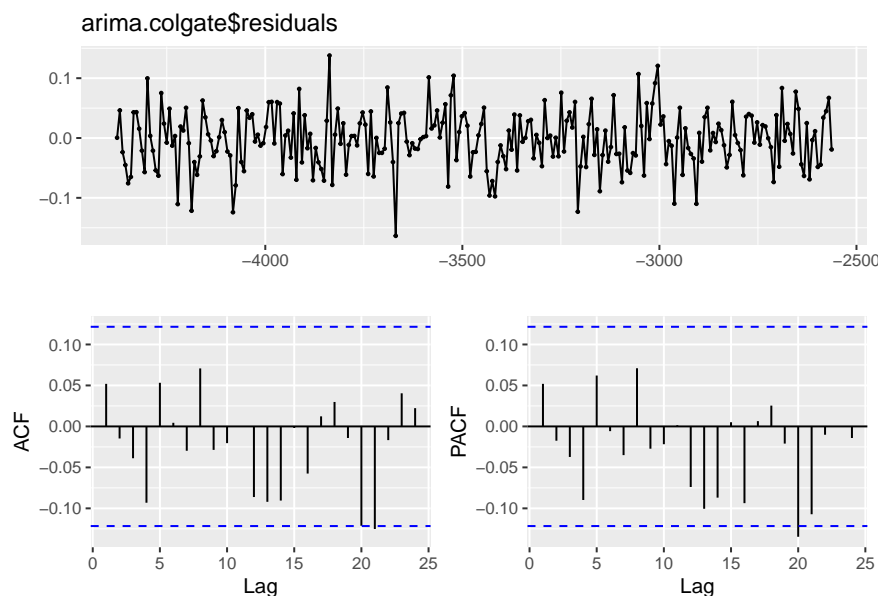
Para Crest vemos que los residuos se comportan como ruido blanco, tanto en ACF como PACF

```
ggtsdisplay(arima.crest$residuals)
```



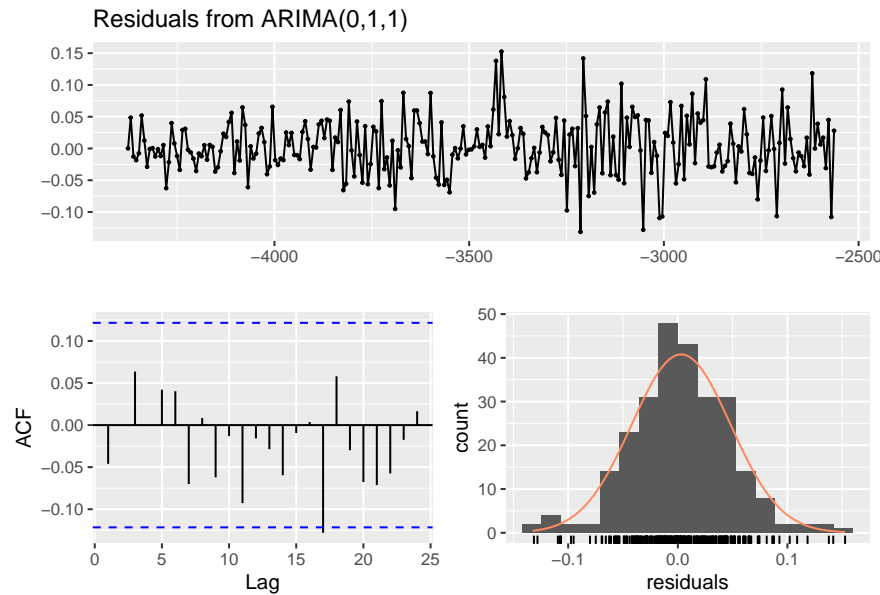
Realizamos el mismo análisis para Colgate y vemos que también son ruido blanco

```
ggtsdisplay(arima.colgate$residuals)
```



Adicionalmente vemos que tanto en Crest como en Colgate los residuos tienen una distribución normal

```
checkresiduals(arima.crest)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)
## Q* = 4.9754, df = 9, p-value = 0.8364
##
## Model df: 1.   Total lags used: 10
```

Si bien con el impacto visual veíamos que los residuos se comportaban como ruido blanco, estadísticamente a través del test Box- Ljung, que no es más que un tipo de prueba estadística de si un grupo de autocorrelaciones de una serie temporal son diferentes de cero. La hipótesis nula asume que existe ruido blanco.

Con un p-valor muy superior al nivel de significación del 5% no podemos rechazar la hipótesis nula, corroborando la información proporcionada por los residuos, comportándose como tales como ruido blanco. Aplicando mayores retardos vemos que el p-valor escada vez mayor.

```
Box.test(arima.crest$residuals,lag = 3, fitdf = 1, type = "Lj")
```

```
##
##  Box-Ljung test
##
## data:  arima.crest$residuals
## X-squared = 1.6314, df = 2, p-value = 0.4423
```

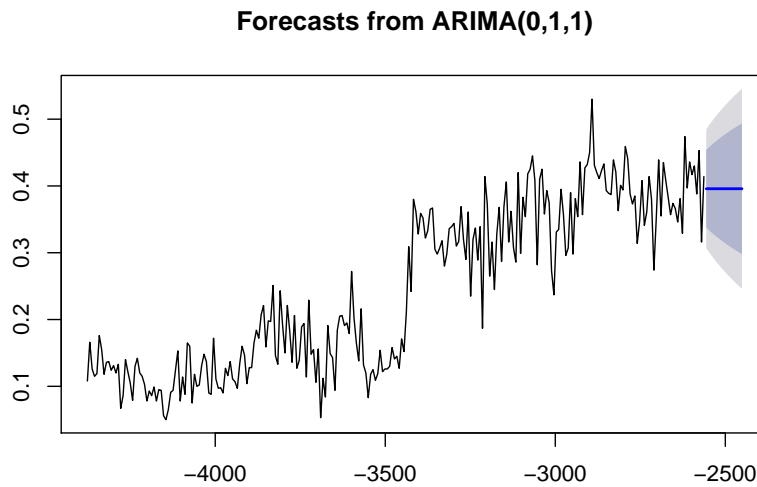
Realizamos el mismo test con Colgate y vemos que exactamente se llegan a las mismas conclusiones

```
Box.test(arima.colgate$residuals,lag = 3, fitdf = 1, type = "Lj")
```

```
##
##  Box-Ljung test
##
## data:  arima.colgate$residuals
## X-squared = 1.1657, df = 2, p-value = 0.5583
```

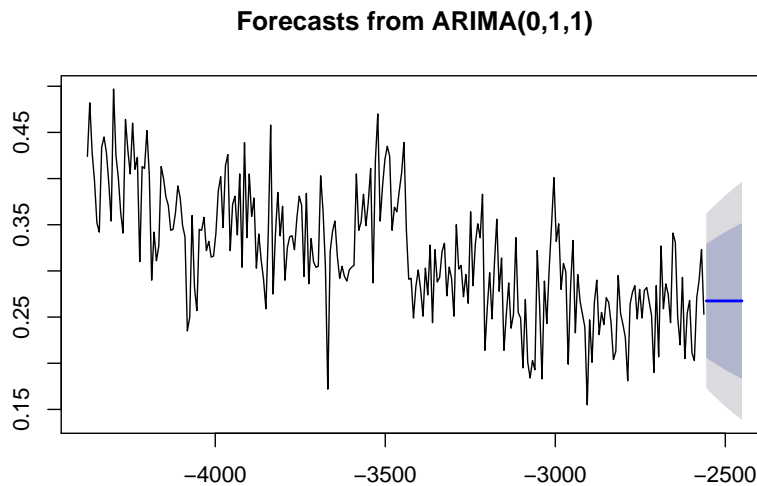
La predicción para el modelo ARIMA (0,1,1) de Crest es

```
fventas.crest = forecast(arima.crest, h = 16)
plot(fventas.crest)
```



La predicción para el modelo ARIMA (0,1,1) de Colgate es

```
fventas.colgate = forecast(arima.colgate, h = 16)
plot(fventas.colgate)
```



DETECCIÓN DE OUTLIERS

Los outliers pueden crear estructuras de autocorrelación espurias, que se evitan añadiendo componentes de intervención. Existen dos tipos de outliers, los aditivos (AO), los cuales afectan a la serie temporal; y los innovativos (IO), los cuales afectan a los residuos.

Para Crest no existen outliers innovativos, sin embargo, sí que existen outliers aditivos en la semana 135, 136

y 138

```
detectA0(arima.crest)
```

```
##           [,1]      [,2]      [,3]
## ind      135.000000 136.000000 138.000000
## lambda2   3.918954  4.372891  4.005427
```

```
detectIO(arima.crest)
```

```
## [1] "No IO detected"
```

Para Colgateo existen tanto outliers innovativos ni outliers aditivos

```
detectA0(arima.colgate)
```

```
## [1] "No A0 detected"
```

```
detectIO(arima.colgate)
```

```
## [1] "No IO detected"
```

ANÁLISIS DE INTERVENCIÓN

La intervención puede deberse al registro de datos erróneos o a observaciones ausentes. En ocasiones, las intervenciones dan lugar a valores atípicos (outliers) que resultan muy improbables dada la historia pasada de la serie temporal. Un modelo de intervención puede utilizarse para:

- Evaluar a posteriori la eficacia de una medida. En nuestro caso el anuncio por parte de la ADA de que Crest era una ayuda importante en cualquier programa de higiene dental.
- Homogeneizar y corregir la correspondiente serie temporal
- Limpiar estructuras de autocorrelación espurias, creadas por los outliers

Para el análisis de intervención de Crest en donde el AIC es -882,72, en donde los coeficientes, tanto de la media móvil como el outlier aditivo a partir de la semana 135, son estadísticamente significativos con ***

```
crest.arimax1 = arimax(oVentasCrest, order = c(0, 1, 1),
                      xtransf = data.frame(primeros = 1*(seq(oVentasCrest) >= 135)),

                      transfer = list(c(0,0)),
                      method = 'ML')

crest.arimax1

##
## Call:
## arimax(x = oVentasCrest, order = c(0, 1, 1), method = "ML", xtransf = data.frame(primeros = 1 *
##      (seq(oVentasCrest) >= 135)), transfer = list(c(0, 0)))
##
## Coefficients:
##      ma1 primeros-MA0
##      -0.7551      0.1477
## s.e.    0.0479      0.0300
##
## sigma^2 estimated as 0.001902:  log likelihood = 443.36,  aic = -882.72
coeftest(crest.arimax1)

##
```

```
## z test of coefficients:
##
##           Estimate Std. Error  z value  Pr(>|z|)
## ma1        -0.755123   0.047854 -15.7797 < 2.2e-16 ***
## primero-MA0 0.147716   0.030035  4.9182 8.736e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para el análisis de intervención de Colgate en donde el AIC es -845,01, en donde los coeficientes, al igual que con los coeficientes de Crest, tanto de la media móvil como el outlier aditivo a partir de la semana 135, son estadísticamente significativos con ***

```
colgate.arimax = arimax(oVentasColgate, order = c(0, 1, 1),
                        xtransf = data.frame(first = 1*(seq(oVentasColgate) >= 135)
                        ),
                        transfer = list(c(0,0)),
                        method = 'ML')#Maxima verosimilitud
```

```
coeftest(colgate.arimax)
```

```
##
## z test of coefficients:
##
##           Estimate Std. Error  z value  Pr(>|z|)
## ma1        -0.804825   0.043637 -18.4437 < 2.2e-16 ***
## first-MA0 -0.101544   0.027859  -3.6449 0.0002675 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FUNCIÓN DE TRANSFERENCIA

La función de transferencia es una herramienta que nos permite analizar cómo se comportará un determinado proceso (anuncio por parte de la ADA), a medida que el tiempo va pasando. La función de transferencia es la forma básica de describir modelos de sistemas lineales basada en la transformación de Laplace, en donde permite obtener la respuesta temporal, la respuesta estática y la respuesta en frecuencia.

Realizamos la función de transferencia con 15 de numerador, vemos que solo dos coeficientes son los estadísticamente significativos

```
#funcion inicial de transferencia
mod0 <- arimax(colgate_134_D,
               order=c(0,1,1),
               include.mean=TRUE,
               xtransf=crest_134_D,
               transfer=list(c(0,15)),
               method="ML")
```

```
coeftest(mod0)
```

```
##
## z test of coefficients:
##
##           Estimate Std. Error  z value  Pr(>|z|)
## ma1        -0.9999995  0.0214756 -46.5645 < 2.2e-16 ***
## T1-MA0     -0.5329365  0.1553428  -3.4307 0.000602 ***
## T1-MA1      0.0075869  0.1852828   0.0409 0.967338
```

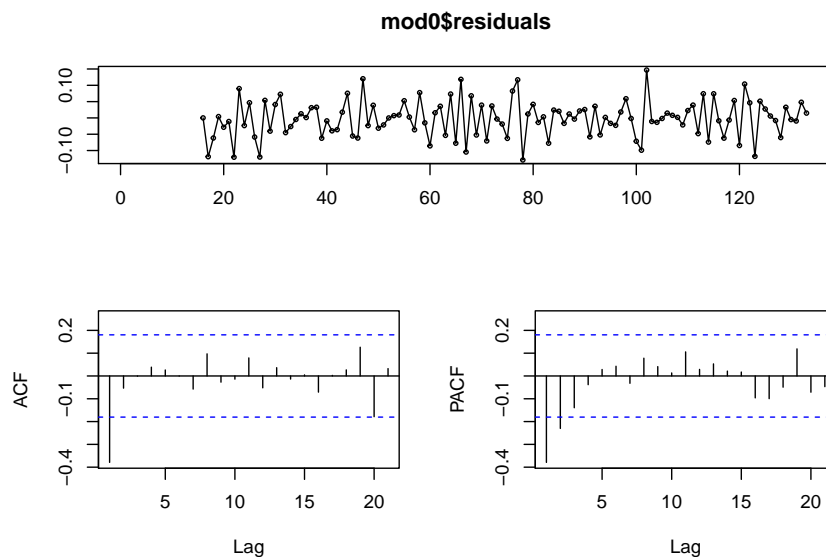


```
## T1-MA2 -0.0431039 0.2016641 -0.2137 0.830749
## T1-MA3 0.1526312 0.2077696 0.7346 0.462573
## T1-MA4 0.0105404 0.2072262 0.0509 0.959434
## T1-MA5 -0.1105735 0.2041136 -0.5417 0.588008
## T1-MA6 0.0390328 0.2036775 0.1916 0.848024
## T1-MA7 -0.1783823 0.2003697 -0.8903 0.373323
## T1-MA8 0.0547611 0.2004141 0.2732 0.784669
## T1-MA9 -0.1667349 0.2022635 -0.8243 0.409743
## T1-MA10 0.0749276 0.2031027 0.3689 0.712191
## T1-MA11 0.2016599 0.2061440 0.9782 0.327952
## T1-MA12 0.0762681 0.2069359 0.3686 0.712456
## T1-MA13 0.0630490 0.2003109 0.3148 0.752947
## T1-MA14 -0.1262427 0.1821569 -0.6930 0.488282
## T1-MA15 -0.0965627 0.1531617 -0.6305 0.528392
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Con nuestro modelo ARIMA (0,1,1) obtenemos un AIC de -304,59 y un RMSE mínimo de 0,0555

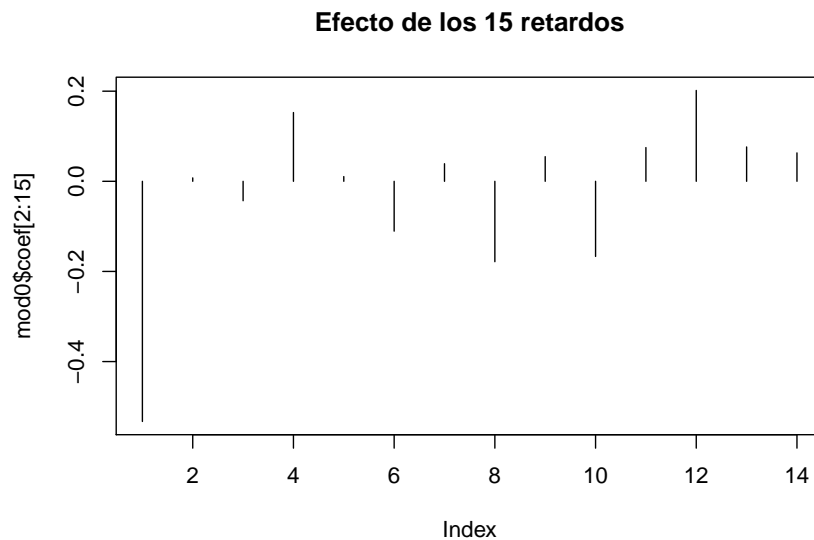
Con respecto a los residuos tanto en ACF como en PACF observamos que sobresalen las franjas azules, por lo que concluimos que los residuos no se comportan como ruido blanco

```
tsdisplay(mod0$residuals)
```



Representando gráficamente la función de transferencia con 15 retardos vemos que el evento acontecido el 1 de Agosto de 1960, tras el anuncio de la ADA, ha sido muy significativo para la semana en particular que provocó una bajada sustancial de la cuota de mercado de Colgate y paralelamente un aumento sustancial de Crest

```
plot(mod0$coef[2:15], type = 'h', main = "Efecto de los 15 retardos")
```



CONCLUSIONES

En base al análisis de intervención podemos asegurar que el gasto en publicidad por parte de nuestro director de Marketing va ha resultar efectivo, siendo especialmente significativo a partir del primer trimestre del año siguiente. Si bien las acciones que se deberían de implementar para aumentar el ROI, captando nuevos clientes no solo en el mercado estadounidense sino también europeo, así como fidelizar a nuestros clientes en la parte noreste de EEUU (Cincinnati y alrededores) son las siguientes (tales medidas se encuentran en el Executive Summary:

1. Lanzamiento un nuevo producto,a nivel local,como el detergente y ver el impacto que tiene en las ventas y en el ROI.
2. Descuentos de más de un 20% por comprar un pack de pasta de dientes Crest e hilo dental en establecimientos comerciales y en supermercados, para ello tendremos ue acordar precios con nuestros proveedores.
3. Pocas empresas son conocidas a nivel internacional a principios de los 60. Darse a conocer a través de medios de comunicación en prensa y en radio son fundamentales y establecer un buen sistema logístico de reparto por todo el mercado de EEUU y Europa es fundamental para el abaratamiento de costes de transporte.

Enlace a GitHub: