# Proyecto-Limabank

Proyecto del curso Temas de Investigación de Operaciones 2020-1

Github: https://github.com/Cubi123/Proyecto-Limabank

## Caso: Agencias

### Problematica

Usted ha sido elegido como asesor de un Gerente de agencia del Banco Limabank. Se requiere la formulacion de una estrategia de planeamiento para el horizonte de tiempo comprendido entre los meses de abril a junio de 2019. Para cumplir con esta tarea se ha recopilado informacion de la agencia desde Enero 2014 hasta Marzo 2019.

### Dataset

* **Arribos Totales** = Clientes que llegan a la agencia.
* **T. Espera Total** = Tiempo promedio que espera el cliente en cola (minutos).
* **T. Atención Total** = Tiempo promedio que demora un cliente con un asesor (minutos).
* **Staff Teórico TM** = Cantidad de asesores en el turno mañana.
* **Staff Teórico TT** = Cantidad de asesores en el turno tarde.
* **Ventas tipo 1** = Cantidad de ventas del producto 1.
* **Ventas tipo 2** = Cantidad de ventas del producto 2.
* **Ventas tipo 3** = Cantidad de ventas del producto 3.
* **Ventas tipo 4** = Cantidad de ventas del producto 4.

### Analisis exploratorio de los datos (EDA)

Data summary

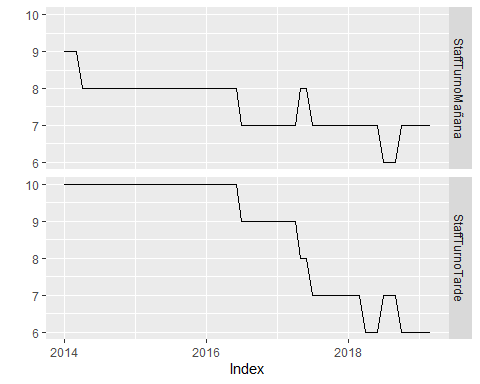
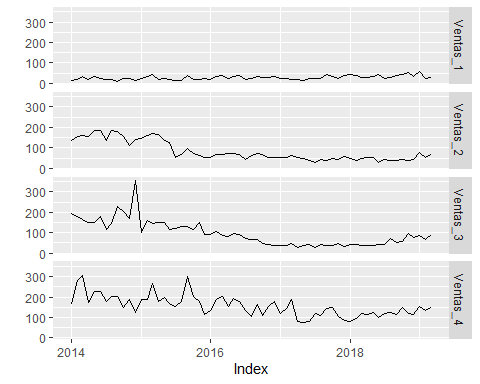
|  |  |
| --- | --- |
| Name | data\_agencia |
| Number of rows | 63 |
| Number of columns | 9 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 9 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

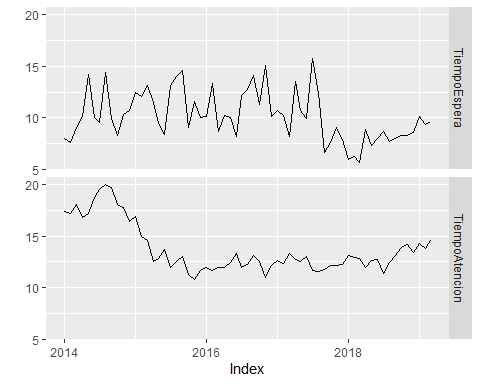
**Variable type: numeric**

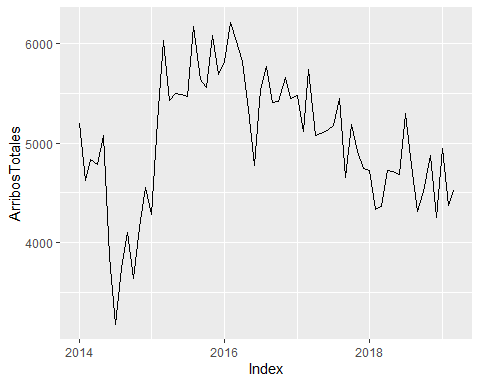
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| ArribosTotales | 2 | 0.97 | 5020.33 | 665.09 | 3176.0 | 4623.0 | 5082.0 | 5485.00 | 6217.0 | ▁▃▇▇▅ |
| StaffTurnoMañana | 0 | 1.00 | 7.51 | 0.67 | 6.0 | 7.0 | 8.0 | 8.00 | 9.0 | ▁▇▁▇▁ |
| StaffTurnoTarde | 0 | 1.00 | 8.63 | 1.57 | 6.0 | 7.0 | 9.0 | 10.00 | 10.0 | ▂▃▁▂▇ |
| TiempoAtencion | 0 | 1.00 | 13.73 | 2.40 | 10.8 | 12.1 | 12.8 | 14.45 | 20.0 | ▇▅▁▂▁ |
| TiempoEspera | 0 | 1.00 | 10.18 | 2.39 | 5.7 | 8.3 | 10.0 | 11.85 | 15.8 | ▃▇▇▃▂ |
| Ventas\_1 | 0 | 1.00 | 27.97 | 10.14 | 10.0 | 20.0 | 26.0 | 34.50 | 57.0 | ▆▇▆▃▁ |
| Ventas\_2 | 1 | 0.98 | 82.66 | 49.09 | 27.0 | 47.0 | 63.0 | 132.50 | 186.0 | ▇▃▁▂▂ |
| Ventas\_3 | 2 | 0.97 | 94.85 | 62.04 | 26.0 | 42.0 | 84.0 | 128.00 | 355.0 | ▇▅▂▁▁ |
| Ventas\_4 | 0 | 1.00 | 156.79 | 51.68 | 78.0 | 119.0 | 151.0 | 187.00 | 305.0 | ▇▇▇▁▂ |
| ### Datos Faltantes |  |  |  |  |  |  |  |  |  |  |

Para completar los datos faltantes, se escogio el metodo de interpolacion. Esto debido a que los datos faltantes son variables exogenas a la empresa.

A continuacion se pueden ver los datos recopilados por la empresa

* Staff por turno tarde y mañana  
  
* Ventas  
  
* Tiempo espera y atencion



* Arribos totales  
  

### Pronostico Arribos totales

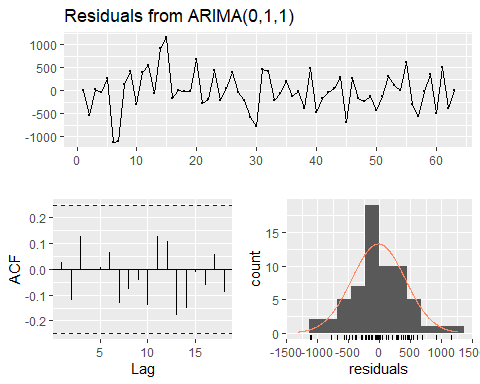
Para realizar un planeamiento correcto es necesario conocer cuantos van a ser los arribos en los siguientes meses, para esto se propone el uso de dos modelos de series de tiempo (Exponential Triple Smoothing y Autorregresive Integrated Moving Average) y comparar sus resultados

**ARIMA**

El ajuste del modelo, nos indica que los datos se ajustan a un modelo ARIMA(0,1,1)

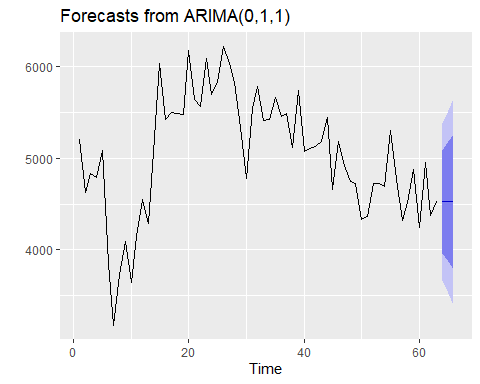
## Series: xts\_agencia$ArribosTotales   
## ARIMA(0,1,1)   
##   
## Coefficients:  
## ma1  
## -0.3816  
## s.e. 0.1210  
##   
## sigma^2 estimated as 188860: log likelihood=-464.16  
## AIC=932.32 AICc=932.52 BIC=936.57  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -16.01653 427.6265 320.1177 -0.9072502 6.681005 0.8655225  
## ACF1  
## Training set 0.02774422

Los errores del modelo cumplen la condicion de normalidad. Tienen un comportamiento de WN (White Noise). El modelo es valido



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(0,1,1)  
## Q\* = 5.8279, df = 9, p-value = 0.757  
##   
## Model df: 1. Total lags used: 10

En el grafico a continuacion se observa el pronostico con intervalos de confianza de 80% y 95%

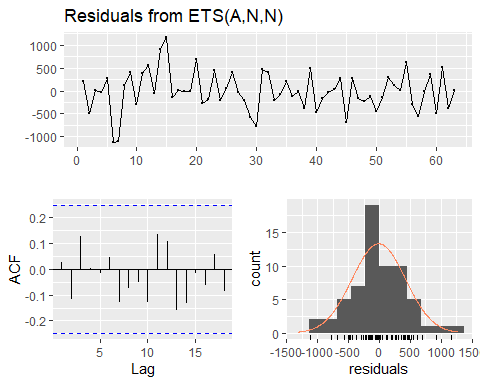


**ETS**

El ajuste del modelo, nos indica que los datos se ajustan a un modelo solo con nivel.

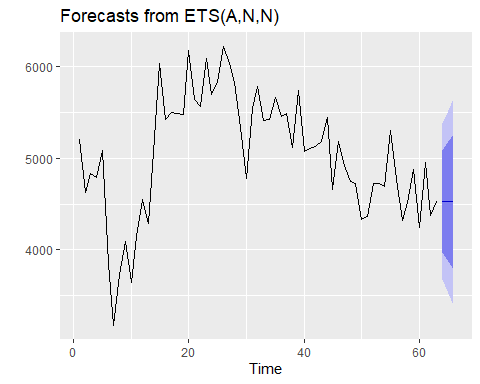
## ETS(A,N,N)   
##   
## Call:  
## ets(y = xts\_agencia$ArribosTotales)   
##   
## Smoothing parameters:  
## alpha = 0.6116   
##   
## Initial states:  
## l = 5002.7653   
##   
## sigma: 434.5695  
##   
## AIC AICc BIC   
## 1030.354 1030.761 1036.783   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -12.4259 427.6159 321.8933 -0.8399322 6.713345 0.8703233  
## ACF1  
## Training set 0.02593694

Los errores del modelo cumplen la condicion de normalidad. Tienen un comportamiento de WN (White Noise). El modelo es valido



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(A,N,N)  
## Q\* = 5.3939, df = 8, p-value = 0.7148  
##   
## Model df: 2. Total lags used: 10

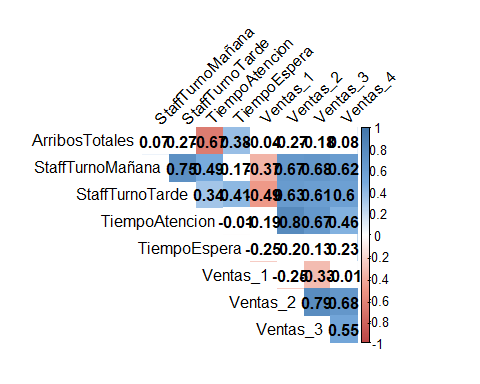
En el grafico a continuacion se observa el pronostico con intervalos de confianza de 80% y 95%



Los valores pronosticados por el ajuste de los modelos ARIMA y ETS son similares, por lo cual es indistinto escoger entre cualquiera de estos modelos. Para el caso de estudio, elegimos el modelo de ARIMA.

### Modelamiento

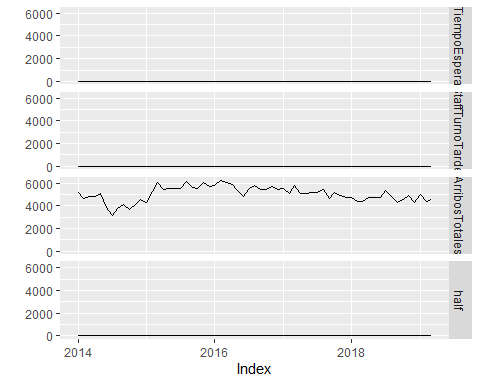
Se tiene como objetivo determinar que variables tienen una mayor influencia en el tiempo de espera actual. Para conocer la relacion entre datos se hace uso del grafico de correlaciones (correlograma).



En el correlograma, se puede observar que las variables con mayor influencia en nuestra variable de impacto, son la cantidad de personas en el staff del turno tarde y los arribos Totales. Para comprobar esto, hacemos uso de una anova. Agregamos datos de la serie de tiempo a nuestro analisis como (mes,año,mitad,trimestre).

## Analysis of Variance Table  
##   
## Response: TiempoEspera  
## Df Sum Sq Mean Sq F value Pr(>F)   
## ArribosTotales 1 51.413 51.413 12.7987 0.0009891 \*\*\*  
## StaffTurnoMañana 1 13.257 13.257 3.3002 0.0773806 .   
## StaffTurnoTarde 1 31.987 31.987 7.9627 0.0076369 \*\*   
## TiempoAtencion 1 21.466 21.466 5.3437 0.0264684 \*   
## Ventas\_1 1 7.683 7.683 1.9127 0.1749585   
## Ventas\_2 1 0.109 0.109 0.0272 0.8700158   
## Ventas\_3 1 2.056 2.056 0.5119 0.4788045   
## Ventas\_4 1 0.324 0.324 0.0806 0.7781299   
## index.num 1 4.433 4.433 1.1036 0.3002893   
## diff 1 13.822 13.822 3.4409 0.0715817 .   
## year 1 7.522 7.522 1.8725 0.1794466   
## year.iso 1 2.203 2.203 0.5483 0.4636713   
## half 1 17.897 17.897 4.4552 0.0416139 \*   
## quarter 1 0.210 0.210 0.0522 0.8205186   
## month 1 3.709 3.709 0.9234 0.3428108   
## wday 1 3.614 3.614 0.8997 0.3490082   
## qday 1 0.022 0.022 0.0055 0.9413965   
## yday 1 1.115 1.115 0.2776 0.6013960   
## mweek 1 0.170 0.170 0.0423 0.8381062   
## week 1 9.216 9.216 2.2942 0.1383579   
## week.iso 1 6.513 6.513 1.6214 0.2108454   
## week2 1 1.925 1.925 0.4792 0.4931137   
## week3 1 1.026 1.026 0.2554 0.6162714   
## week4 1 0.015 0.015 0.0038 0.9511583   
## Residuals 37 148.630 4.017   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

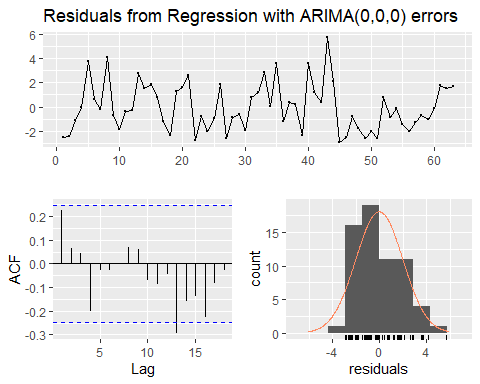
De acuerdo al analisis de varianza realizada, las variables con mayor significacion en el modelo son StaffTurnoTarde, TotalArribos, half (1:primera mitad del año, 2:segunda mitad del año) y TiempoAtencion.



**ARIMA**

Ajustamos el modelo de regresion dinamica, modelando los errores con un modelo de arima.

## Series: data\_arima$TiempoEspera   
## Regression with ARIMA(0,0,0) errors   
##   
## Coefficients:  
## ArribosTotales StaffTurnoTarde half  
## 9e-04 0.4904 1.0325  
## s.e. 3e-04 0.1520 0.4603  
##   
## sigma^2 estimated as 4.128: log likelihood=-132.52  
## AIC=273.03 AICc=273.72 BIC=281.61  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.01960623 1.982773 1.633191 -4.157514 16.62299 0.8333976  
## ACF1  
## Training set 0.2276741



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(0,0,0) errors  
## Q\* = 7.8426, df = 7, p-value = 0.3467  
##   
## Model df: 3. Total lags used: 10

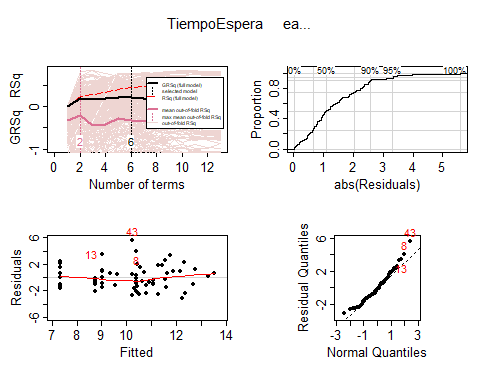
Dado que el modelo resultante es un ARIMA(0,0,0), se opta por usar un modelo lineal.

**MARS**  
Debido a que el modelo de ARIMA solo explica un error normal (White noise) se opta por usar un modelo Multivariado adaptativo de regresion por splines, con el objetivo de conocer el impacto de las variables en el modelo.

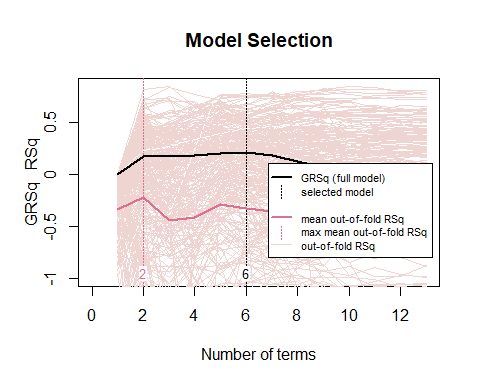
Se divide el modelo en 80% y 20% y se entrena el modelo utilizando cross validation, 10 folds

## Call: earth(formula=TiempoEspera~ArribosTotales+StaffTurnoTarde+half,  
## data=data, keepxy=TRUE, nfold=10, ncross=30, varmod.method="lm")  
##   
## coefficients  
## (Intercept) 5.9502256  
## half 1.3863387  
## h(ArribosTotales-4837) 0.0200871  
## h(ArribosTotales-5077) -0.0838682  
## h(ArribosTotales-5130) 0.0646695  
## h(StaffTurnoTarde-7) 0.5607839  
##   
## Selected 6 of 13 terms, and 3 of 3 predictors  
## Termination condition: Reached nk 21  
## Importance: ArribosTotales, StaffTurnoTarde, half  
## Number of terms at each degree of interaction: 1 5 (additive model)  
## GCV 4.549264 RSS 195.2573 GRSq 0.2184975 RSq 0.4502646 CVRSq -0.5024413  
##   
## Note: the cross-validation sd's below are standard deviations across folds  
##   
## Cross validation: nterms 3.52 sd 1.36 nvars 1.97 sd 0.90  
##   
## CVRSq sd MaxErr sd  
## -0.502 1.32 8.37 4.02  
##   
## varmod: method "lm" min.sd 0.219 iter.rsq 0.011  
##   
## stddev of predictions:  
## coefficients iter.stderr iter.stderr%  
## (Intercept) 1.30271780 1.05932 81  
## TiempoEspera 0.08680348 0.104799 121  
##   
## mean smallest largest ratio  
## 95% prediction interval 8.571845 7.602924 9.707235 1.276777  
##   
## 68% 80% 90% 95%   
## response values in prediction interval 79 92 95 98

No se detecta presencia de herestocedasticidad en nuestro modelo. Algunos outlier en cuanto a los errores residuales. El modelo es valido



Observamos que el numero de bases optimas para la estimacion de nuestro modelo es 6, en funcion del GRSq

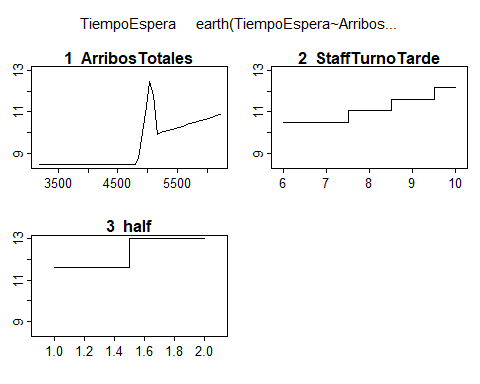


Observamos que las variables mas importantes para la estimacion de nuestro modelos en orden de mayor a menor importante son los arribos totales, la cantidad de personas en el staff turno tarde y si nos encontramos en la primera o segunda mitad del año.

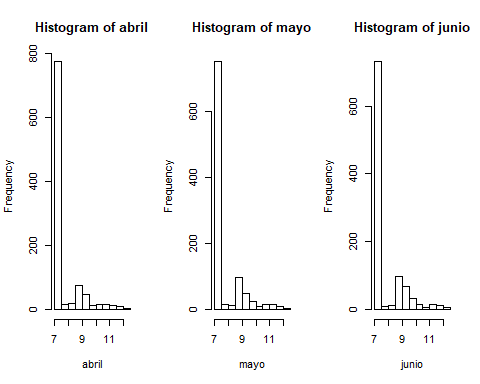
## nsubsets gcv rss  
## ArribosTotales 5 100.0 100.0  
## StaffTurnoTarde 3 45.5 62.1  
## half 2 38.3 50.3

En el siguiente grafico, observamos el impacto que tiene la variacion de las variables en nuestra variable respuesta Tiempo Espera. En la primera mitad del año, el tiempo de espera es menor que en la segunda mitad del año. Asimismo, un aumento en la cantidad de personas en el staff del turno tarde, se ve reflejado en un aumento en el tiempo espera. Por ultimo, se observa un impacto mas pronunciado en la variable respuesta, en el rango de 4837 y 5077. Apartir de 5130 se equilibra el impacto de esta variable.

## plotmo grid: ArribosTotales StaffTurnoTarde half  
## 5103.066 9 1

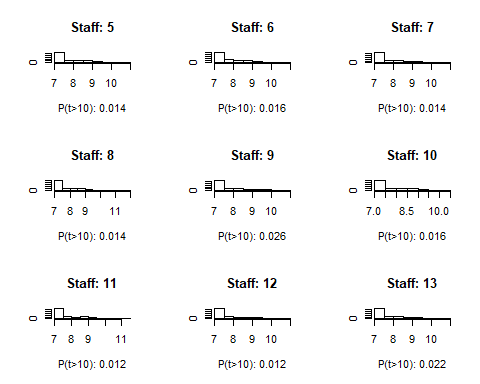


Por ultimo, para pronosticar los tiempos de espera para Abril, Mayo, Junio , se corrieron 1000 simulaciones, utilizando los valores generados por nuestro ajuste por ARIMA, numero de personas staff tarde igual a 6 (mismo que los meses anteriores) y half igual a 1, ya que nos encontramos en la primera mitad de año



A traves de los histogramas generados para cada mes, observamos que se cumple con el objetivo de la gerencia de mantener en 10min el tiempo limite de espera. La probabilidad de superar ese tiempo de espera para Abril, Mayo y Junio es de 0.049,0.071,0.063

Luego de hacer el analisis de sensibilidad variando el numero de staff de turno tarde entre 5 y 13 obtenemos los siguientes resultados



Observamos que la mejor opcion es mantener o reducir el staff a 5