

# Tarea Minería de Datos - Análisis y Predicción de Series Temporales

Gerard Chicot Navalls

28/7/2022

```
library(readxl)
library(ggplot2)
library(knitr)
library(forecast)
library(zoo)
```

1. Introducción: Presentación de la serie a analizar.

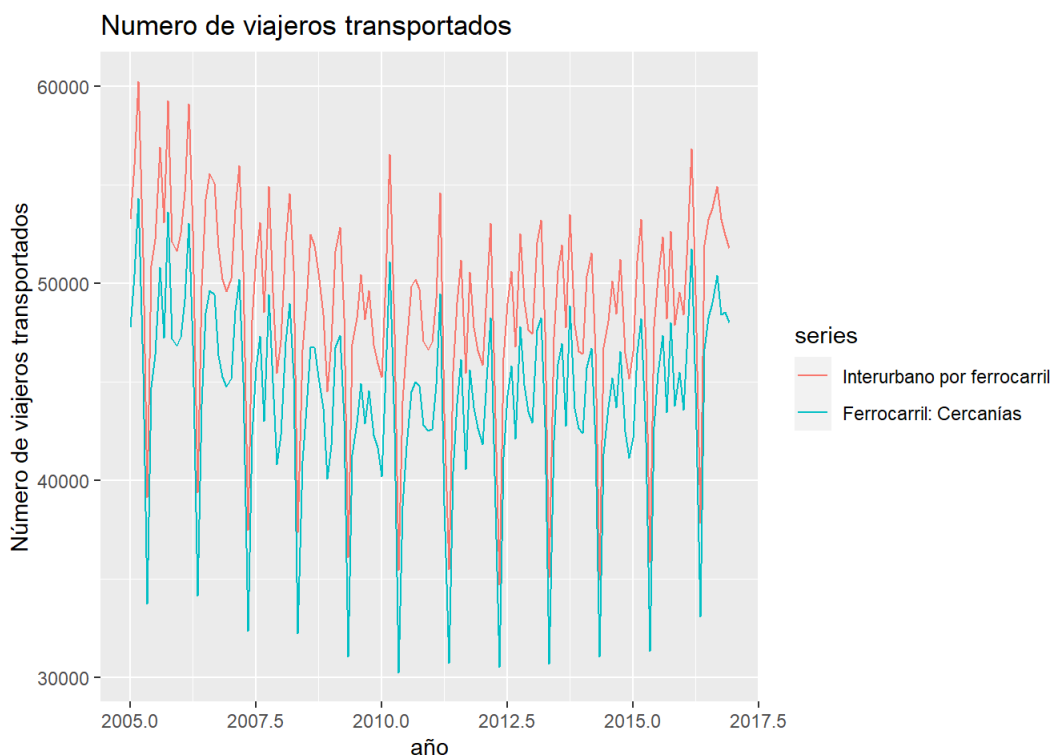
```
datos <- read_xlsx("Interurbano ferrocarril.xlsx")
datos
```

```
## # A tibble: 144 x 3
##   Periodo `Interurbano por ferrocarril` `Ferrocarril: Cercanías`
##   <chr>          <dbl>          <dbl>
## 1 2019M12          53279          47789
## 2 2019M11          56035          50461
## 3 2019M10          60265          54299
## 4 2019M09          51356          45453
## 5 2019M08          39167          33748
## 6 2019M07          50909          44711
## 7 2019M06          52253          46405
## 8 2019M05          56923          50814
## 9 2019M04          53108          47231
## 10 2019M03          59260          53618
## # ... with 134 more rows
```

2. Representación gráfica y descomposición estacional (si tuviera comportamiento estacional)

```
ferrocarril <- ts(datos[,1], start= c(2005,1), frequency=12)
ferrocarril_INT <- ts(datos[,2], start= c(2005,1), frequency=12)
ferrocarril_CERC <- ts(datos[,3], start= c(2005,1), frequency=12)

autoplot(ferrocarril) + ggtitle("Numero de viajeros transportados") + xlab("año") + ylab("Número de viajeros transportados")
```

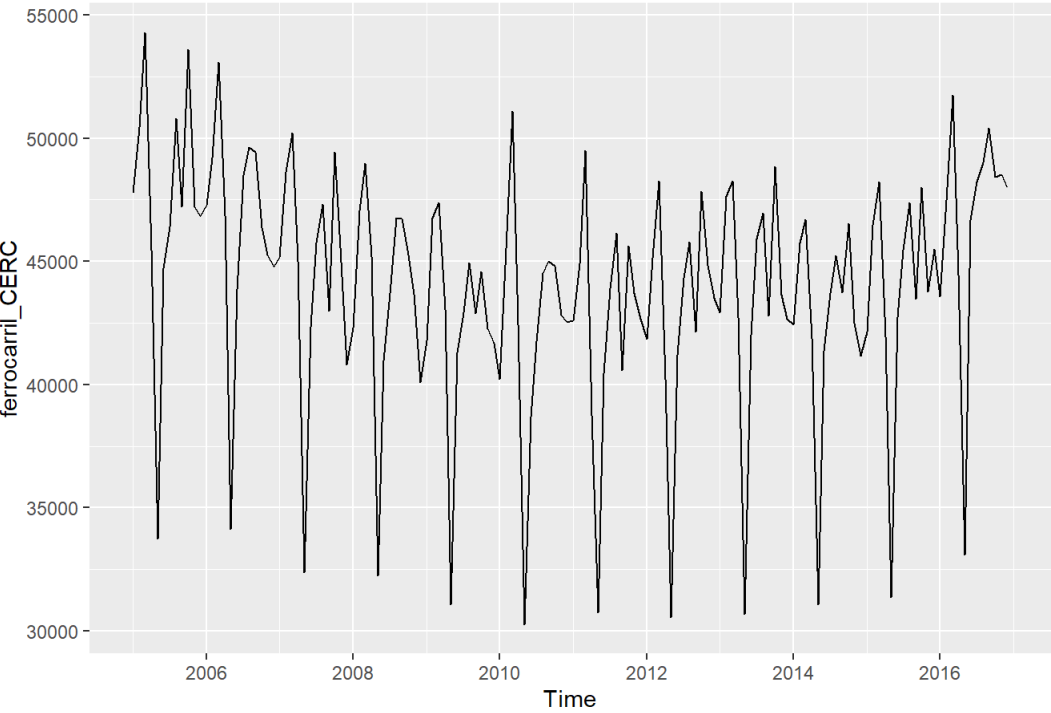


La serie es estacionaria, vemos como la media es constante.

Vemos como hay una pequeña tendencia decreciente hasta el año 2010 y luego empieza a crecer poco a poco.

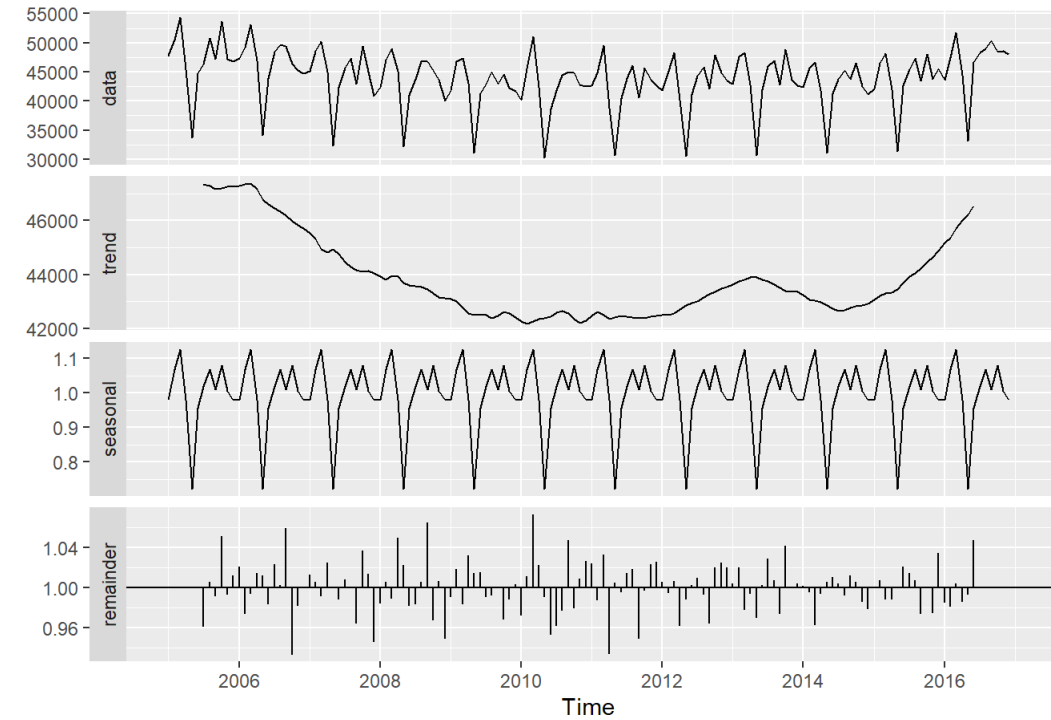
Si que tiene un comportamiento estacional cada 12 meses.

```
autoplot(ferrocarril_CERC)
```



```
ferrocarril_CERC_COMP <- decompose((ferrocarril_CERC), type=c("multiplicative"))
autoplot(ferrocarril_CERC_COMP, ts.colour = "blue")
```

Decomposition of multiplicative time series

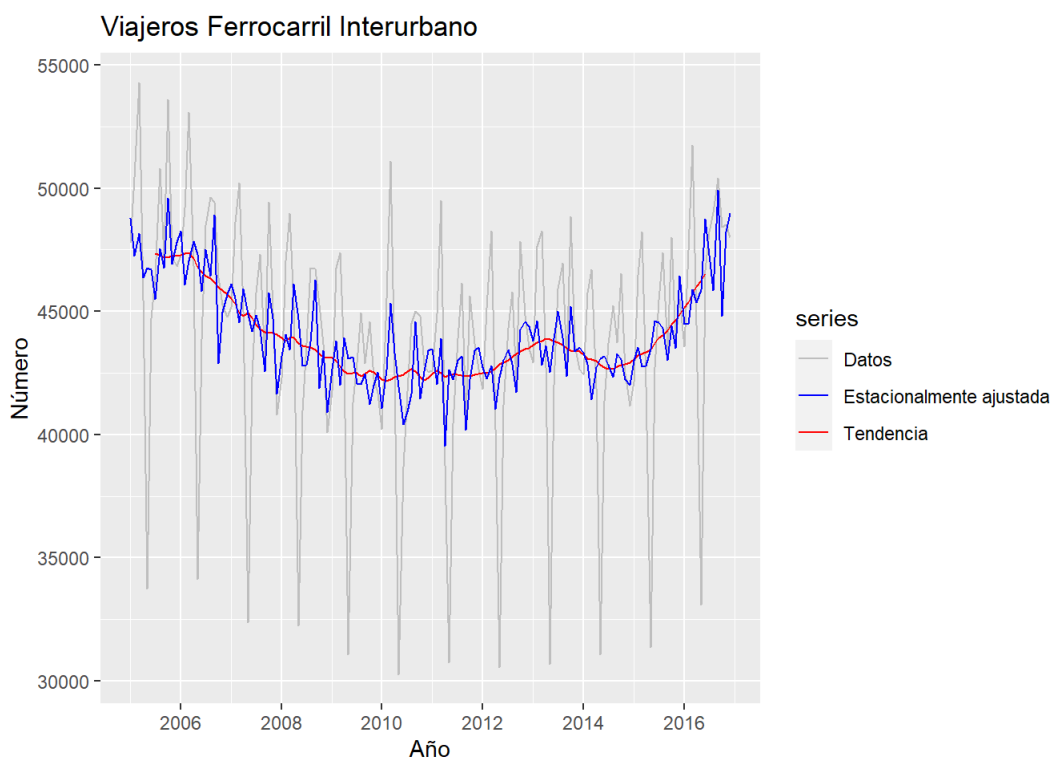


```
knitr::kable(ferrocarril_CERC_COMP$figure, digits =2,caption = "Coef Estacionalidad")
```

Coef Estacionalidad

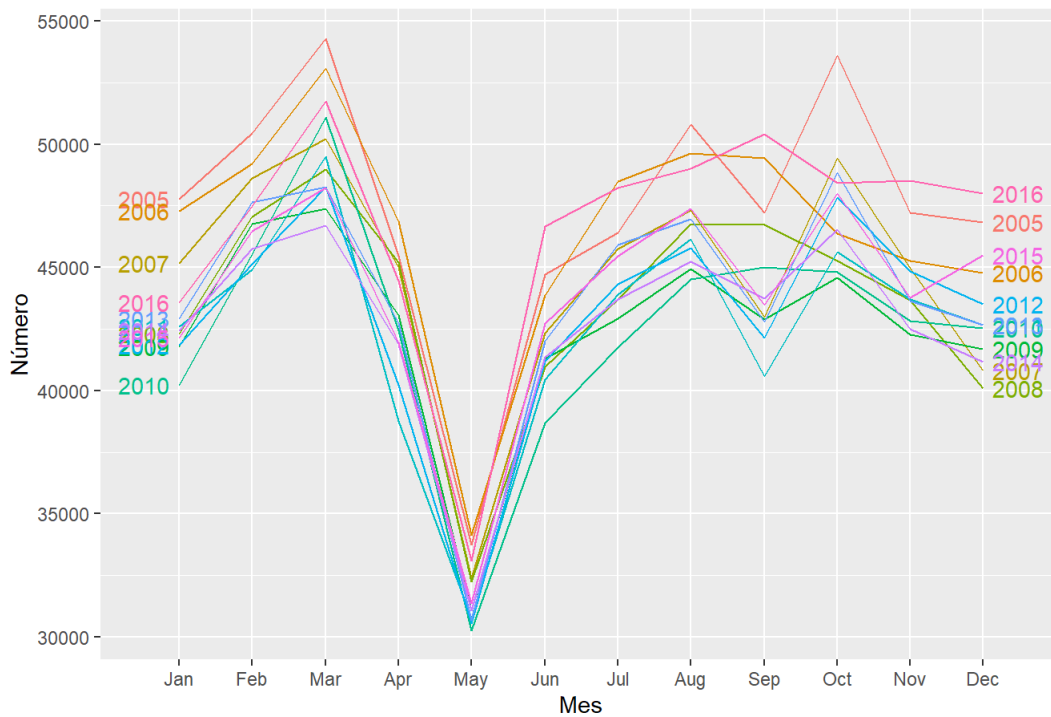
	x
	0.98
	1.07
	1.13
	0.98
	0.72

```
autoplot(ferrocarril_CERC, series="Datos") +  
  autolayer(trendcycle(ferrocarril_CERC_COMP), series="Tendencia") +  
  autolayer(seasadj(ferrocarril_CERC_COMP), series="Estacionalmente ajustada") +  
  xlab("Año") + ylab("Número") +  
  ggtitle("Viajeros Ferrocarril Interurbano") +  
  scale_colour_manual(values=c("gray","blue","red"),  
  
  breaks=c("Datos", "Estacionalmente ajustada", "Tendencia"))
```



```
ggseasonplot(ferrocarril_CERC, year.labels=TRUE, year.labels.left=TRUE) + xlab("Mes")+  
  ylab("Número") + ggtitle("Viajeros Ferrocarril Interurbano")
```

## Viajeros Ferrocarril Interurbano



Podemos observar los coeficientes de estacionalidad donde el valor más grande corresponde al Marzo con un 1.13. Esto significa que hay un 13% más de viajeros en el ferrocarril de Cercanías que en la media del año. El coeficiente menor es en Mayo con un valor de 0.72, esto significa que en ese mes hay un 28% menos de viajeros que en la media del año.

Como podemos ver en la representación gráfica de los valores de la serie, todos los años tienen la misma estructura y se ve también que va aumentando el número de turistas cada año.

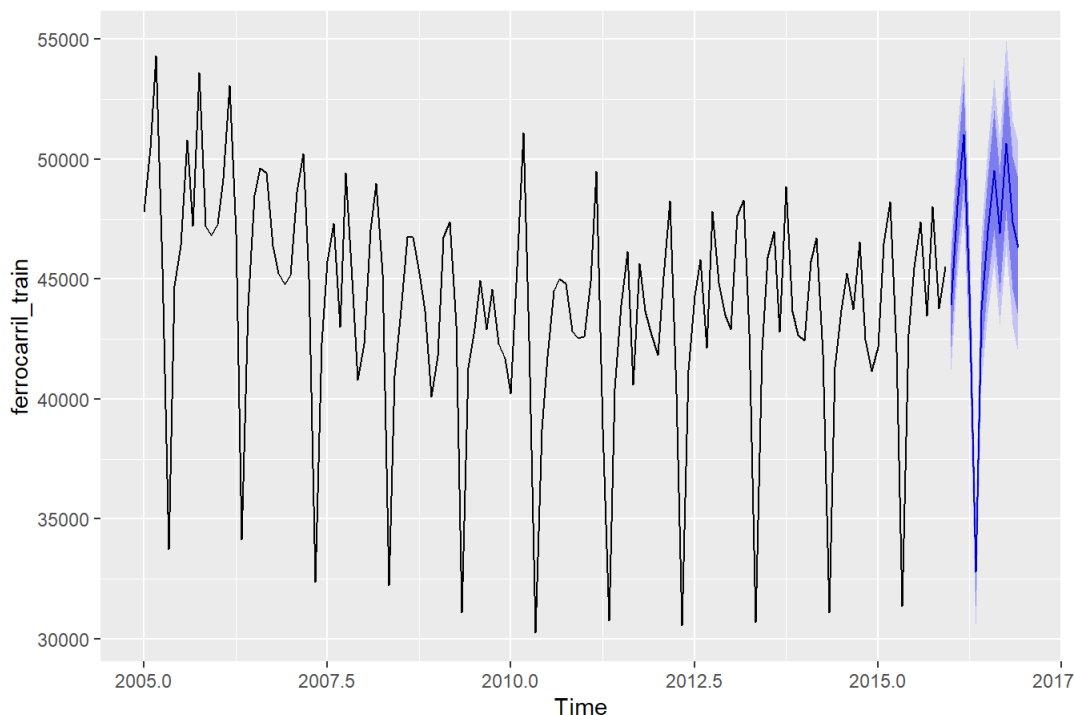
- Para comprobar la eficacia de los métodos de predicción que vamos a hacer en los siguientes apartados reservamos los últimos datos observados (un periodo en las series estacionales o aproximadamente 10 observaciones) para comparar con las predicciones realizadas por cada uno de los métodos. Luego ajustamos los modelos sobre la serie sin esos últimos datos en los siguientes apartados

```
ferrocarril_train <- window(ferrocarril_CERC, end=c(2015,12))
```

- Encontrar el modelo de suavizado exponencial más adecuado, mostrando una tabla con los estimadores de los parámetros del modelo elegido. Para dicho modelo, representar gráficamente la serie observada y la suavizada con las predicciones para un periodo que se considere adecuado. Mostrar una tabla con las predicciones

```
fit_ferrocarril_CERC <- hw(ferrocarril_train, h=12, seasonal="multiplicative", level = c(80,95))
autoplot(fit_ferrocarril_CERC)
```

### Forecasts from Holt-Winters' multiplicative method



knitr::kable(fit\_ferrocarril\_CERC, digits =4,caption = "Predicciones ")

Predicciones

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2016	43893.62	42154.54	45632.71	41233.92	46553.32
Feb 2016	48091.89	46152.57	50031.20	45125.96	51057.82
Mar 2016	51033.81	48927.38	53140.23	47812.31	54255.31
Apr 2016	44407.23	42520.36	46294.10	41521.51	47292.95
May 2016	32796.11	31353.65	34238.58	30590.06	35002.17
Jun 2016	43719.92	41719.40	45720.44	40660.39	46779.46
Jul 2016	47021.69	44773.44	49269.95	43583.28	50460.10
Aug 2016	49544.37	47060.11	52028.62	45745.02	53343.71
Sep 2016	46904.86	44431.40	49378.33	43122.02	50687.70
Oct 2016	50654.56	47839.46	53469.66	46349.23	54959.88
Nov 2016	47344.79	44568.02	50121.55	43098.10	51591.48
Dec 2016	46319.41	43450.18	49188.64	41931.30	50707.51

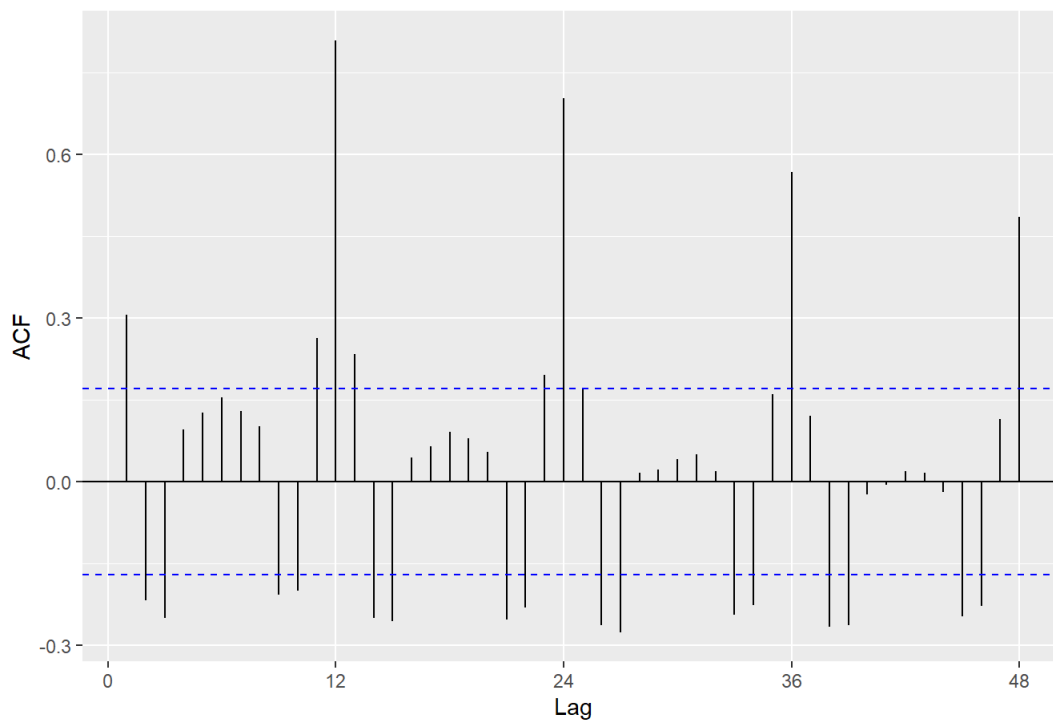
fit\_ferrocarril\_CERC[["model"]]

```
## Holt-Winters' multiplicative method
##
## Call:
## hw(y = ferrocarril_train, h = 12, seasonal = "multiplicative",
##
## Call:
##   level = c(80, 95))
##
## Smoothing parameters:
##   alpha = 0.1563
##   beta  = 0.034
##   gamma = 1e-04
##
## Initial states:
##   l = 47889.6625
##   b = -74.4093
##   s = 0.9806 1.0072 1.0829 1.0077 1.0696 1.0202
##       0.9533 0.7187 0.9781 1.1298 1.0701 0.9817
##
## sigma: 0.0309
##
##   AIC   AICc   BIC
## 2563.694 2569.062 2612.701
```

5. Representar la serie y los correlogramas. Decidir que modelo puede ser ajustado. Ajustar el modelo adecuado comprobando que sus residuales están incorrelados. (Sintaxis, tablas de los parámetros estimados y gráficos)

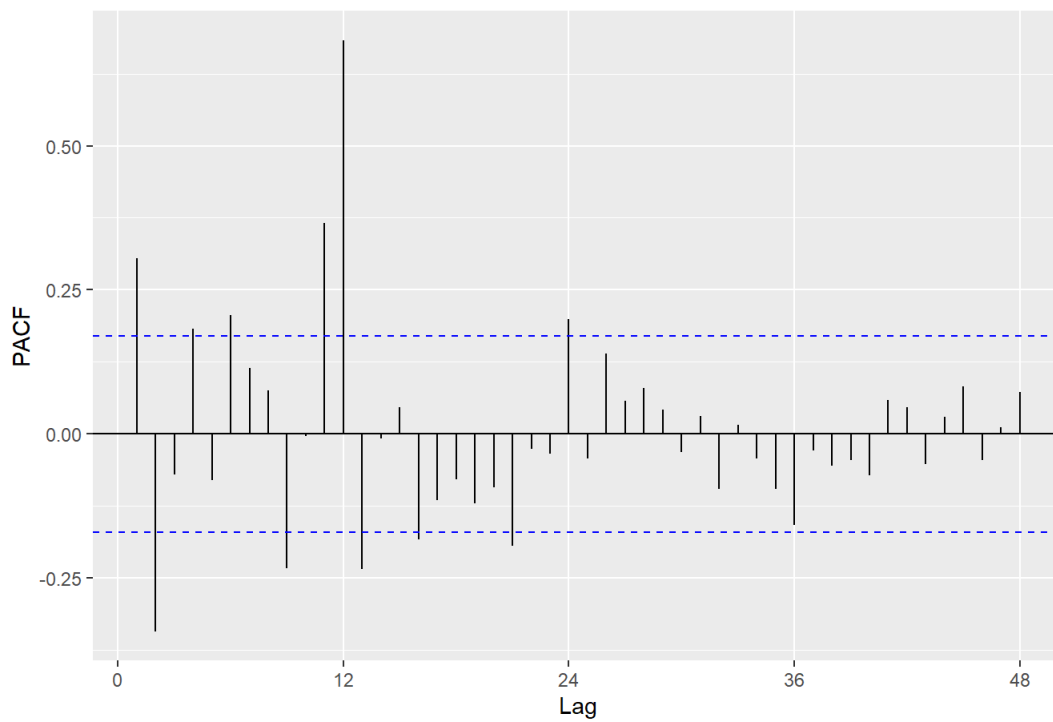
ggAcf(ferrocarril\_train, lag=48)

Series: ferrocarril\_train



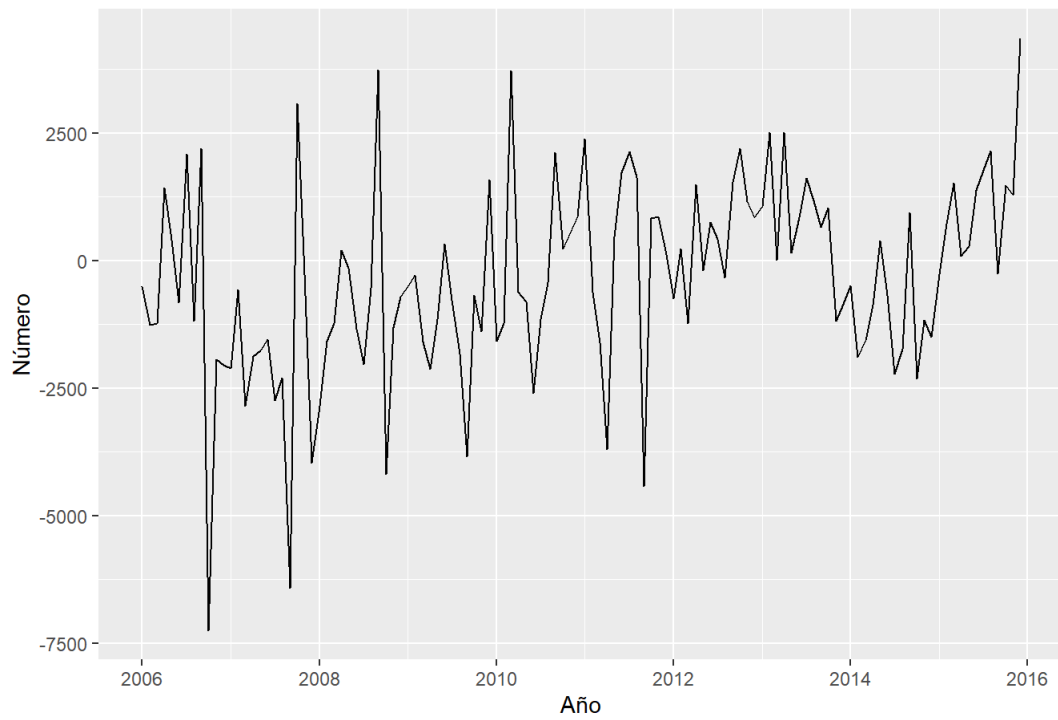
```
ggPacf(ferrocarril_train, lag=48)
```

Series: ferrocarril\_train

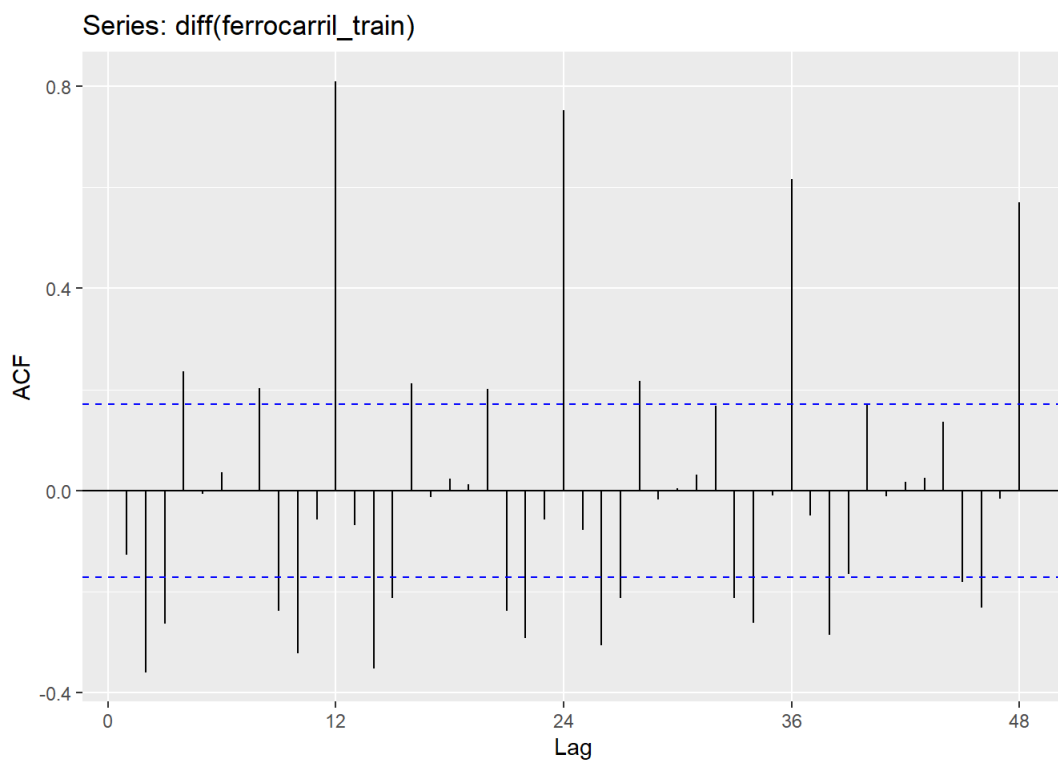


```
autoplot(diff(ferrocarril_train,12))+ ggtitle("Viajeros ferrocarril") + xlab("Año") +  
ylab("Número")
```

## Viajeros ferrocarril

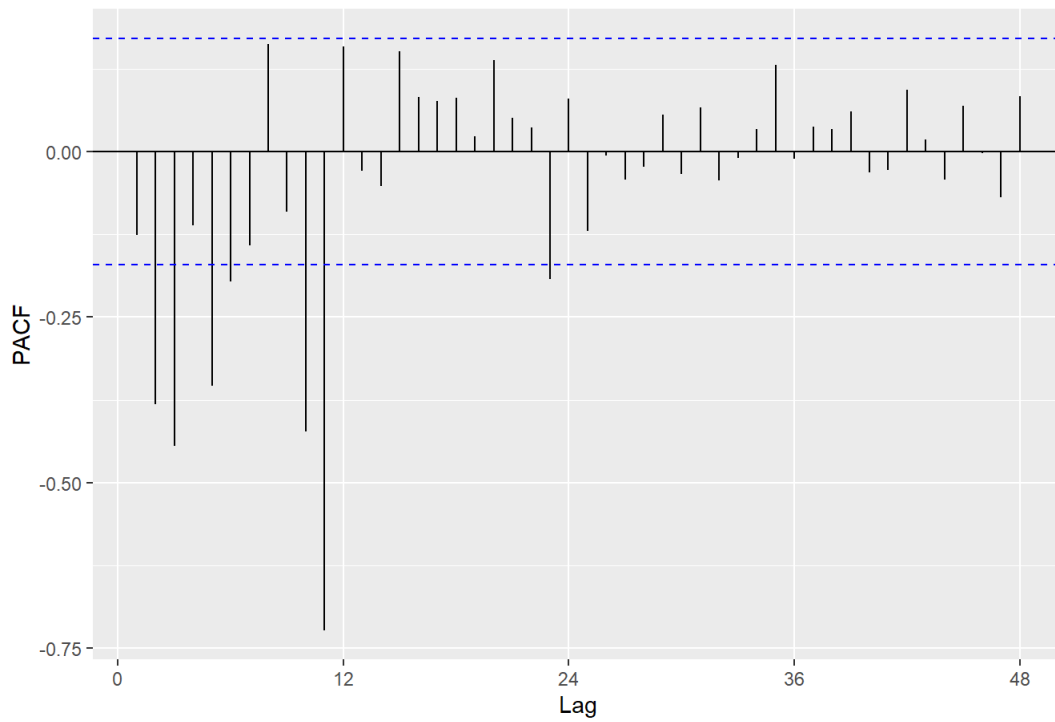


```
ggAcf(diff(ferrocarril_train), lag=48)
```

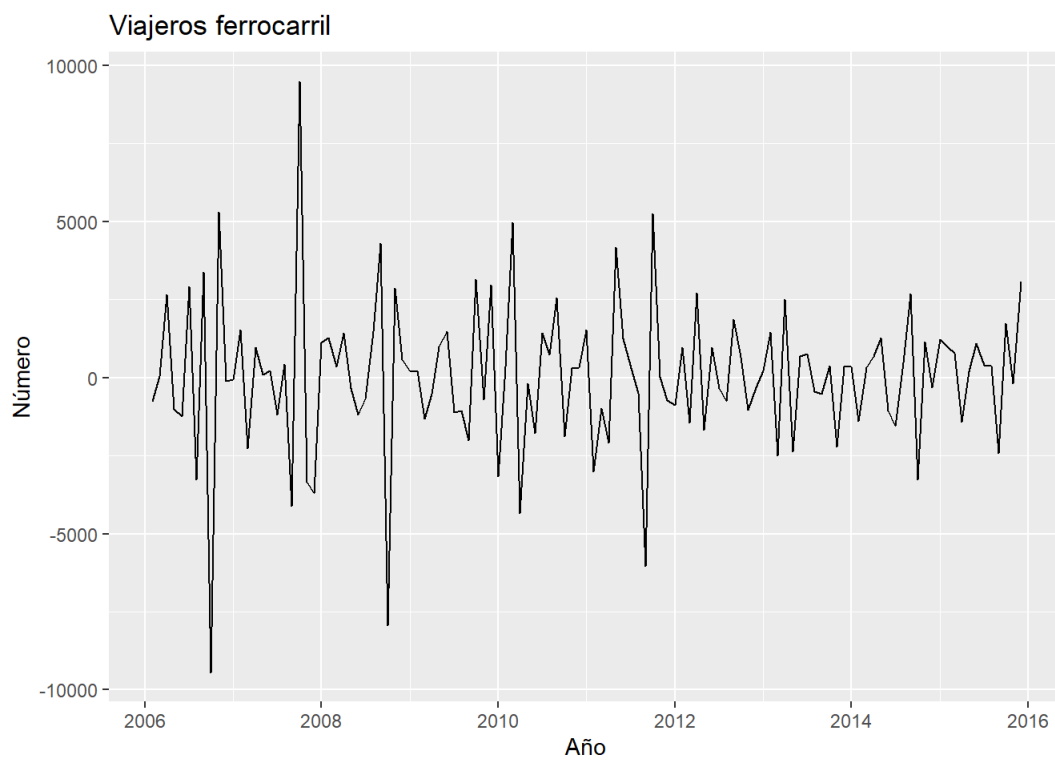


```
ggPacf(diff(ferrocarril_train), lag=48)
```

Series: diff(ferrocarril\_train)



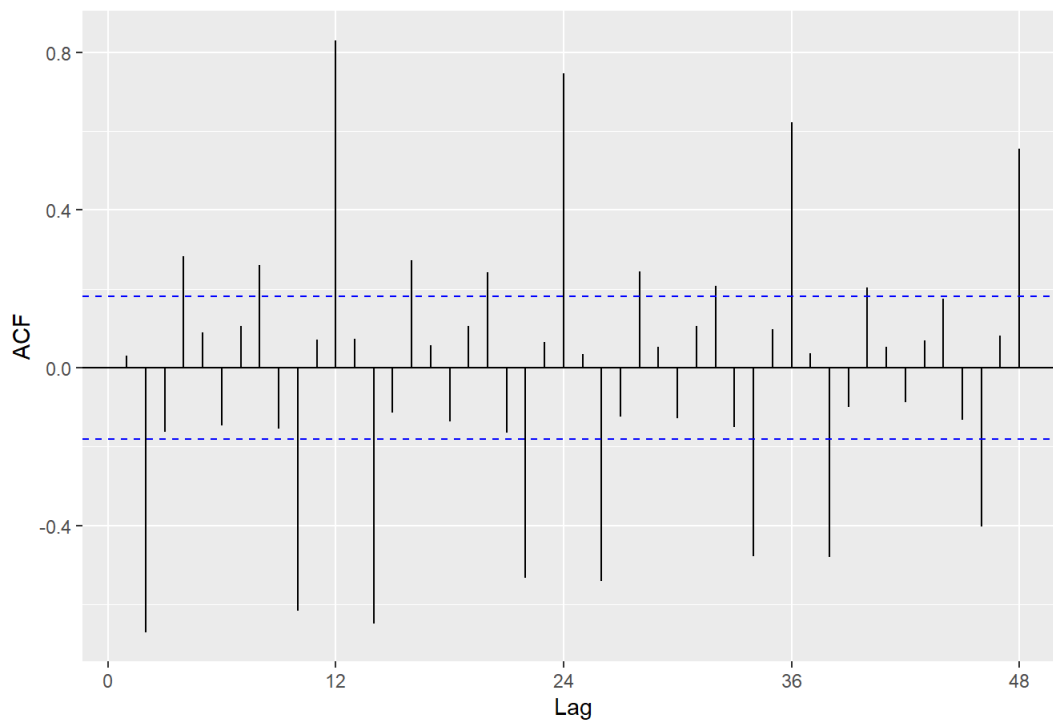
```
autoplot(diff(diff(ferrocarril_train),12))+ ggtitle("Viajeros ferrocarril") + xlab("Año") + ylab("Número")
```



```
ggAcf(diff(diff(ferrocarril_train),14), lag=48)
```

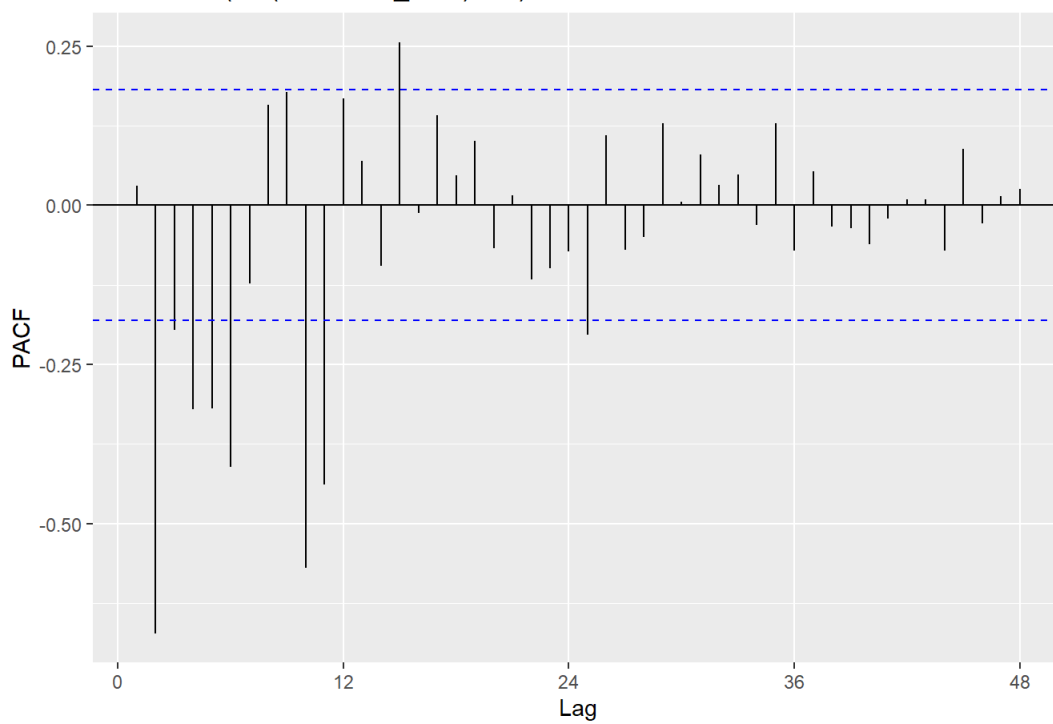


Series: diff(diff(ferrocarri1\_train), 14)



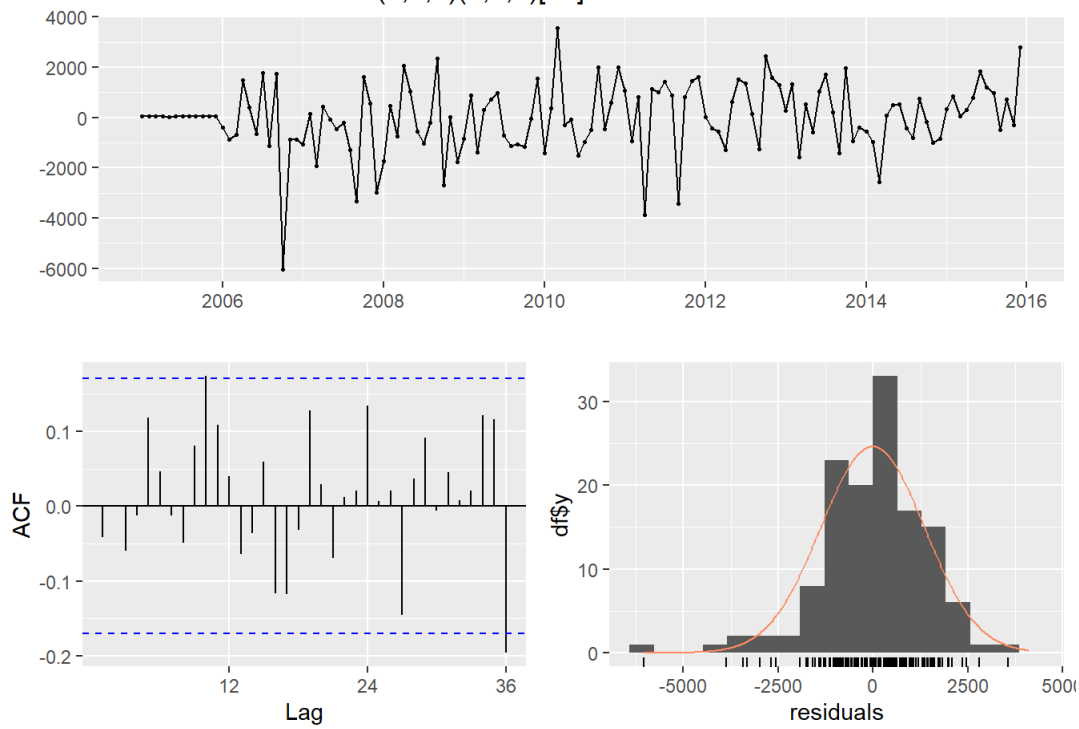
```
ggPacf(diff(diff(ferrocarri1_train),14), lag=48)
```

Series: diff(diff(ferrocarri1\_train), 14)



```
fitferrocarri1 <- Arima((ferrocarri1_train),c(1,0,1),seasonal=c(0,1,1))  
checkresiduals(fitferrocarri1)
```

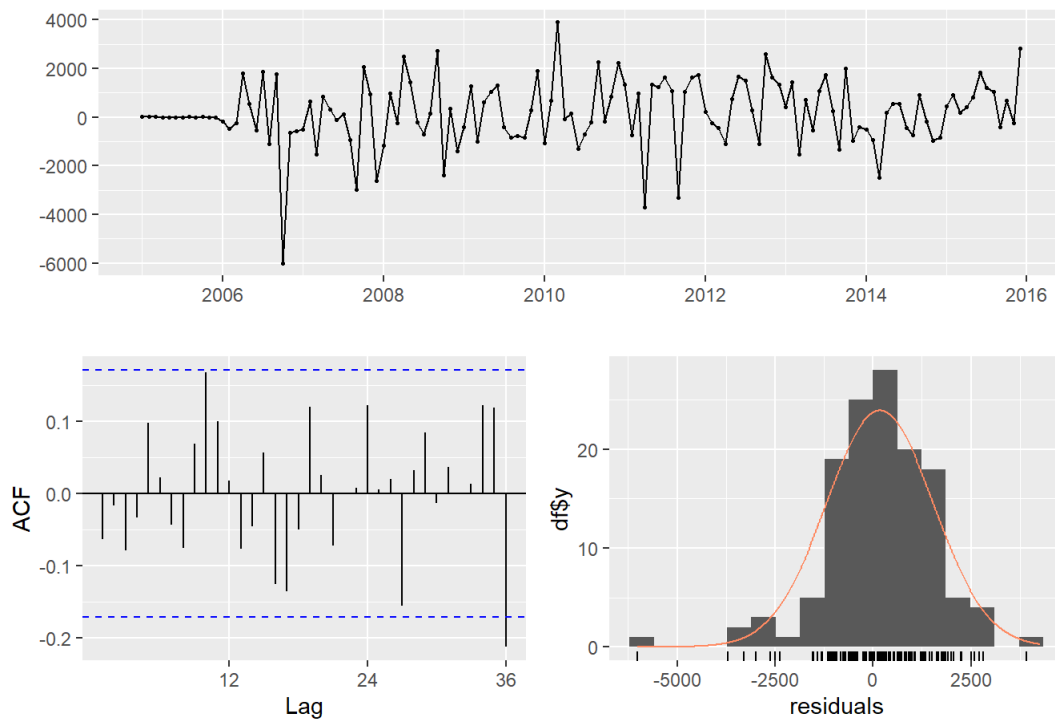
Residuals from ARIMA(1,0,1)(0,1,1)[12]



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(1,0,1)(0,1,1)[12]
## Q* = 22.858, df = 21, p-value = 0.3516
##
## Model df: 3. Total lags used: 24
```

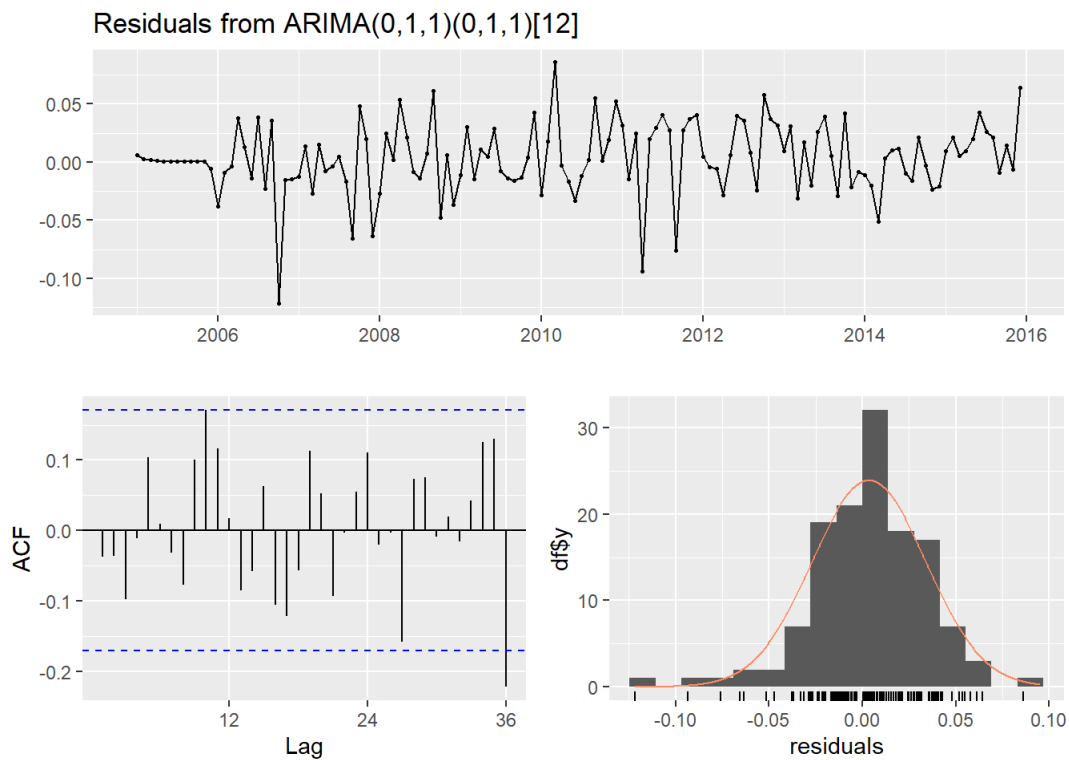
```
fitferrocarril2 <- auto.arima(ferrocarril_train)
checkresiduals(fitferrocarril2)
```

Residuals from ARIMA(0,1,1)(0,1,1)[12]



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,1,1)(0,1,1)[12]
## Q* = 23.235, df = 22, p-value = 0.3885
##
## Model df: 2. Total lags used: 24
```

```
fitferrocarril3 <- auto.arima(log(ferrocarril_train))
checkresiduals(fitferrocarril3)
```



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,1,1)(0,1,1)[12]
## Q* = 24.944, df = 22, p-value = 0.2998
##
## Model df: 2. Total lags used: 24
```

```
knitr::kable(accuracy(fitferrocarril3), digits = 4, caption = "Medias de ajuste")
```

Medias de ajuste

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.0034	0.0306	0.0227	0.0318	0.2123	0.671	-0.0377

Vemos como todos los modelos están incorrelados y finalmente me he decidido por el modelo número 3. Es decir, explica que toda la dependencia que hay en el modelo gracias a que el p-value es mayor a 0.05.

6. Escribir la expresión algebraica del modelo ajustado con los parámetros estimados.

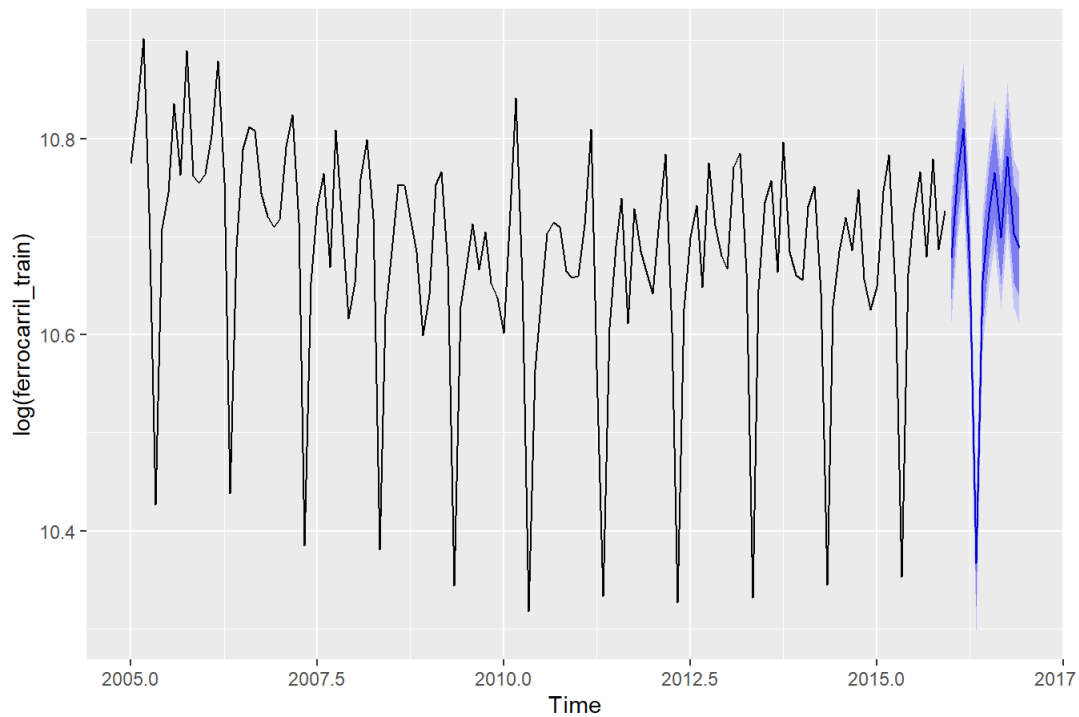
```
fitferrocarril3
```

```
## Series: log(ferrocarril_train)
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##      ma1      sma1
##    -0.8002 -0.7750
## s.e.  0.0511  0.1039
##
## sigma^2 = 0.001056: log likelihood = 234.63
## AIC=-463.26 AICc=-463.05 BIC=-454.92
```

7. Calcular las predicciones y los intervalos de confianza para las unidades de tiempo que se considere oportuno, dependiendo de la serie, siguientes al último valor observado. Representarlas gráficamente.

```
predi3 <- forecast(fitferrocarril3,h=12)
autoplot(predi3)
```

## Forecasts from ARIMA(0,1,1)(0,1,1)[12]



8. Comparar las predicciones obtenidas con cada uno de los métodos (suavizado y ARIMA) con los valores observados que habíamos reservado antes. Conclusiones.

```
autoplot(ferrocarril_CERC) +
  autolayer(forecast(fit_ferrocarril_CERC,h=12), series="suabizado", PI=FALSE) +
  autolayer(forecast(fitferrocarril1,h=12), series="manual", PI=FALSE) +
  autolayer(forecast(fitferrocarril2,h=12), series="automatico", PI=FALSE)+
  autolayer(exp(predi3$mean), series="Logaritmico")+
  ggtitle("Predicciones por diferentes modelos ") + xlab("Año") + ylab("Número") +
  guides(colour=guide_legend(title="Forecast"))
```

