

DATA624 - Exercise 3.8

Harpreet Shoker

For your retail time series (from Exercise 3 in Section 2.10):

Loading the retail data

```
retaildata <- readxl::read_excel("retail.xlsx", skip=1)
myts <- ts(retaildata[, "A3349873A"],
  frequency=12, start=c(1982,4))
```

a) Split the data into two parts using

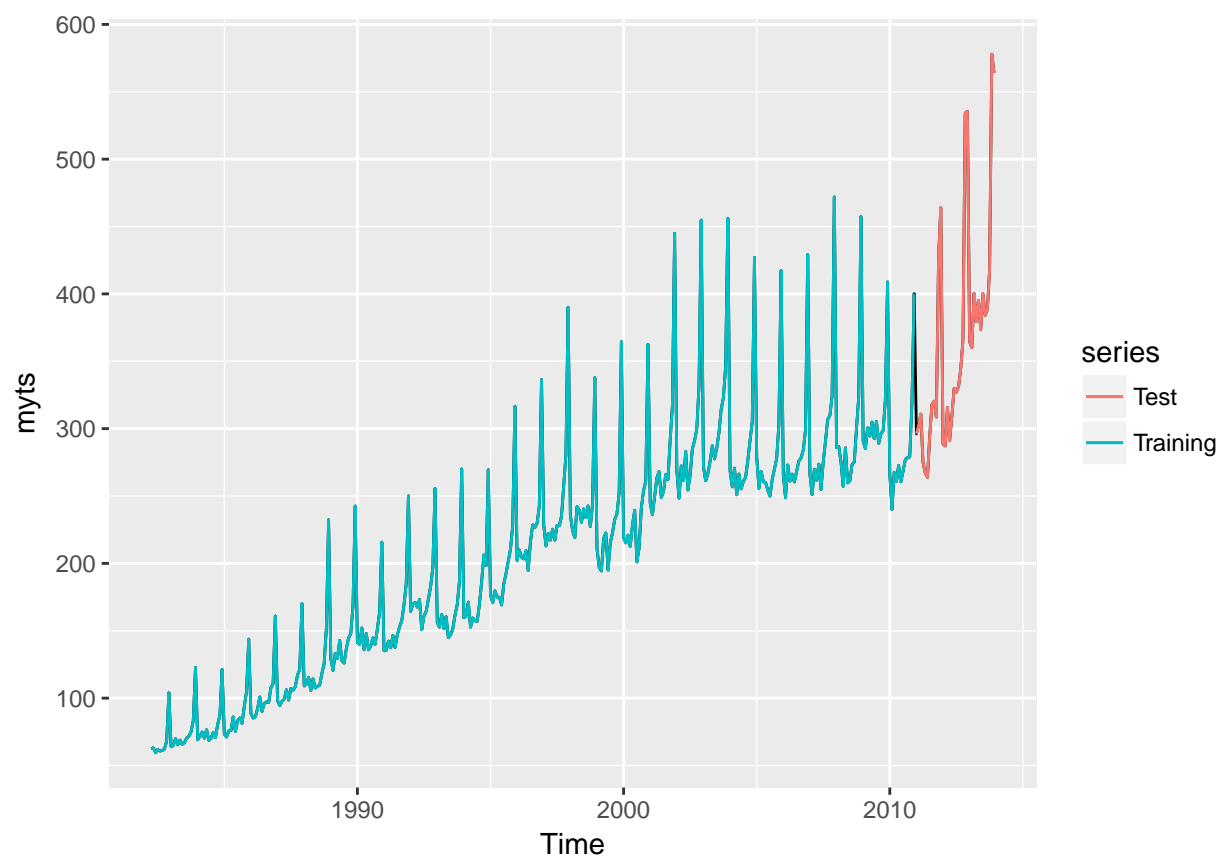
```
myts.train <- window(myts, end=c(2010,12))
myts.test <- window(myts, start=2011)
```

b) Check that your data have been split appropriately by producing the following plot.

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 3.4.4
```

```
autoplot(myts) + autolayer(myts.train, series="Training") +
  autolayer(myts.test, series="Test")
```



c) Calculate forecasts using `snaive` applied to `myts.train`.

```
fc <- snaive(myts.train)
```

d) Compare the accuracy of your forecasts against the actual values stored in `myts.test`.

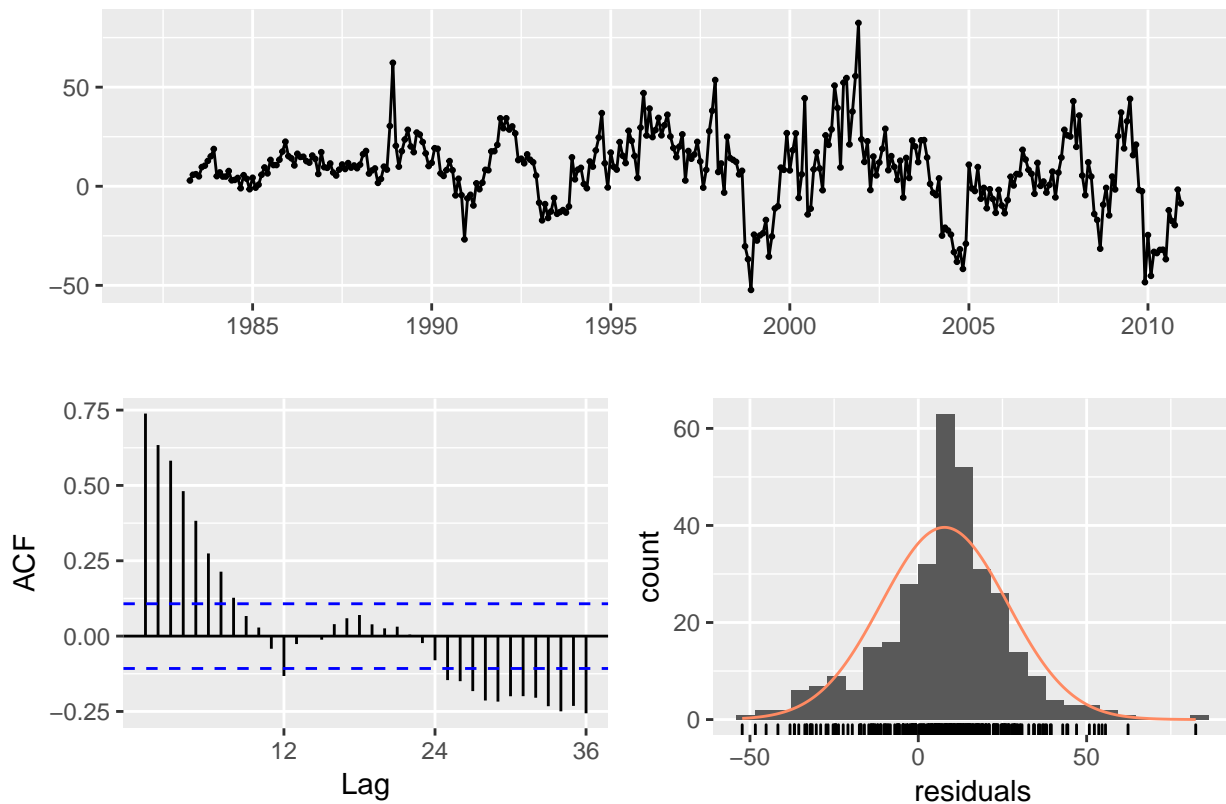
```
accuracy(fc,myts.test)
```

```
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set  7.772973 20.24576 15.95676  4.702754  8.109777  1.000000
## Test set     55.300000 71.44309 55.78333 14.900996 15.082019  3.495907
##               ACF1 Theil's U
## Training set  0.7385090      NA
## Test set      0.5315239  1.297866
```

e) Check the residuals. Do the residuals appear to be uncorrelated and normally distributed?

```
checkresiduals(fc)
```

Residuals from Seasonal naive method



```
##
## Ljung-Box test
##
## data: Residuals from Seasonal naive method
## Q* = 624.45, df = 24, p-value < 2.2e-16
##
## Model df: 0. Total lags used: 24
```

From the above it seems like residuals are correlated to each other. Residuals are not normally distributed.

f) How sensitive are the accuracy measures to the training/test split?

```
myts2.train <- window(myts, end=c(2011,12))
myts2.test <- window(myts, start=2012)
fc2 <- snaive(myts2.train)
accuracy(fc2,myts.test)
```

```
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set  8.828696 21.81237 16.76145  4.922591  8.223701  1.0000
## Test set     65.429167 78.69376 67.34583 15.853615 16.520064  4.0179
##               ACF1 Theil's U
## Training set  0.7198310      NA
## Test set     0.7328968  1.356191
```

The accuracy measures are sensitive to the training/test split. Here we changed the train/test split percentage and run the accuracy check again and that results in low values in accuracy measure indicators. Comparing this to original matrix clearly indicates that the measures are sensitive to the split.