Week 4 Assignment – 50 Points

1. (25 Points) We use “pneu\_flu.rda” timeseries dataset
   1. Partition dataset into train and validation. Make a reasonable choice for partitions’ size.

**Training set and Validation set:**

A black and white text

Description automatically generated with medium confidence

* 1. Use the regression model to fit a forecasting model. Run the model on validation.

**Model Summary:**

A screenshot of a computer

Description automatically generated

* 1. write this models in its mathematical form.

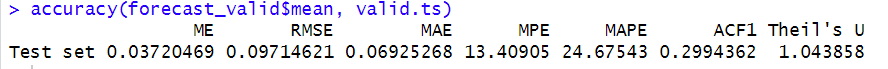
F(t) = -0.384514+(-0.001126\*trend) +(-0.107397\*season2) + (-0.228725\*season3) + (-0.290717\*season4) +(-0.331887\*season5)+(-0.343136\*season6)+(-0.334444\*season7)+(-0.343126\*season8)+ (-0.339717\*season9) +(-0.310769\*season10) +( -0.288005\*season11) + (-0.189023\*season12)+ €

* 1. Put training, time series trend, actual validation, forecasted validation on the same plot. Interpret the result through your visualizations.

A graph of a graph

Description automatically generated with medium confidence

* **Trend Line:** The trend line is not a straight line, indicating that there is a non-linear trend in the data. This suggests that the data is not following a constant rate of change over time, but rather shows some curvature or variation in the trend.
* **Forecasted Validation:** The forecasted validation appears to have minimal errors, suggesting that the forecasting models are performing reasonably well in capturing the underlying patterns and trends in the data.
  1. Get the performance of this model.



* 1. Forecast for the next 4 month in the future.

A number of numbers and letters

Description automatically generated with medium confidence

* 1. Plot the dataset and the 3 months forecast in one plot. What can you say for your forecast errors? There will be no errors until we have the actual values(original data) of that

A graph with lines and numbers

Description automatically generated

* 1. What is the forecasting model you use for part f ? can you evaluate this forecasting model? Explain.
* I have used **linear regression model** (tslm) with trend and seasonal components. This model is specified using the formula PIDeaths ~ trend + season and it is estimated with a Box-Cox transformation (lambda = 1) to ensure that the residuals are approximately normally distributed.
* To evaluate this forecasting model, we can assess its performance using various accuracy measures such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and others.
* Based on the accuracy metrics and visual inspection of the forecasted values, we can determine the effectiveness of the forecasting model. If the model produces accurate forecasts that closely match the observed data, it indicates that the model is performing well.
  1. Use the forecasting model to manually forecast for April and July 1979. Write the detail of your work.

**With linear trend and additive seasonality**

Note: if we are forecasting an observation in a specific season say, M*i ,* only this season is one and the rest are all zeros

**For April 1979:**

M=4, t =132+4 =136

-0.0007828(-0.2430176 + 0

= - 0.7853713

**For July 1979:**

M=7, t =132+7 =139

-0.0007828(-0.2949156

= - 0.8396177

1. (25 Points) Use air passengers count in “AirPassenger.csv” file to:
   1. Built timeseries of this data. Choose proper seasonality frequency.

Seasonality Frequency =12

* 1. Put the timeseries and its level on the same plot.

A graph showing the flight passengers

Description automatically generated with medium confidence

* 1. Put the timeseries and its trend and level on the same plot.

Make sure your trend is accurately identified.

A graph with lines and numbers

Description automatically generated

* 1. Partition data in 24 months validation and rest for training

A calendar with numbers and letters

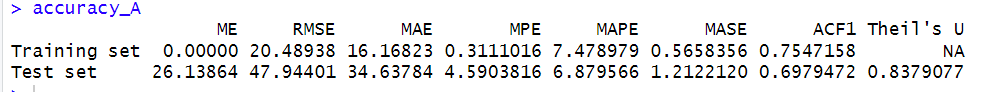
Description automatically generated

* 1. Fit the regression model on the training to build a forecasting model (model A).

A screenshot of a computer

Description automatically generated

* 1. Run the model (model A) on validation and get the RMSE, MPE, and MAPE



**RMSE = 47.94401**

**MPE = 4.5903816**

**MAPE = 6.879566**

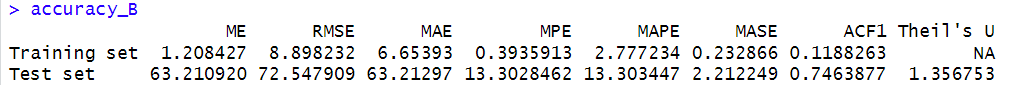
* 1. Use the appropriate model setting for Exponential Smoothing function ets() to build a forecast model (model B).

Ets() function will choose the appropriate model automatically, here it choose model = (M,Ad,M)

A screenshot of a computer

Description automatically generated

* 1. Run the model (model B) on validation and get the RMSE, MPE, and MAPE



**RMSE = 72.547909**

**MPE = 13.3028462**

**MAPE = 13.303447**

* 1. Get the accuracy forecast and compare the RMSE, MAE, and MAPE of these models with your benchmark model. Which one is a better model
* From the below result of accuracy on Test set, **Model\_A** has lower RMSE , MAE, MAPE hence Model\_A is better than Model\_B and benchmark model (Seasonal Naïve)

A screenshot of a computer

Description automatically generated

* 1. Put training, validation and the validation forecasts of A and B models on the same plot

A graph showing a line of growth

Description automatically generated with medium confidence

* 1. Use model A to manually calculate the air passenger count for November and December of 1961. Do not use R.

I have used **linear regression model** (tslm) for AirPassengers dataset and result shown below

**A screenshot of a computer

Description automatically generated**

**For November 1961:**

**Given time series has t =144 time periods**

M=11, t =144+11 =155

(-35.51996

= 439.33913

**For December 1961:**

M=12, t =144+12 =156

(-9.18029)

= 468.33913