8.3 Time Series

- Time Series is a series of data collected over time.

- It's data collected at specific time points.

e.g. Analysis of website checks.
fuel Price
Volume of sales at Retail Outlet
Stock Price

- Data to be collected consistently and regularly and then only can be ~~[illegible]~~ used for forecasting.

We ~~sa~~ have sales data for a particular year and we want to predict sales for next 6 monts.

Regression Modelling	Time Series Forecasting
• Requires explanatory variables	• Requires sales figures
• To predict the sales of 2017 we need variables like Price, Location etc.	• By plotting the sales data of previous year (based on 1 variable) sales can be predicted.

When we have only 1 variable, the most likely method or choice of forecast would be Time Series Forecasting.

Time Series models are built on the premise that

- The future will mimic similar pattern as the past
- Information needed to generate the forecast is contained in the Time Series of the data

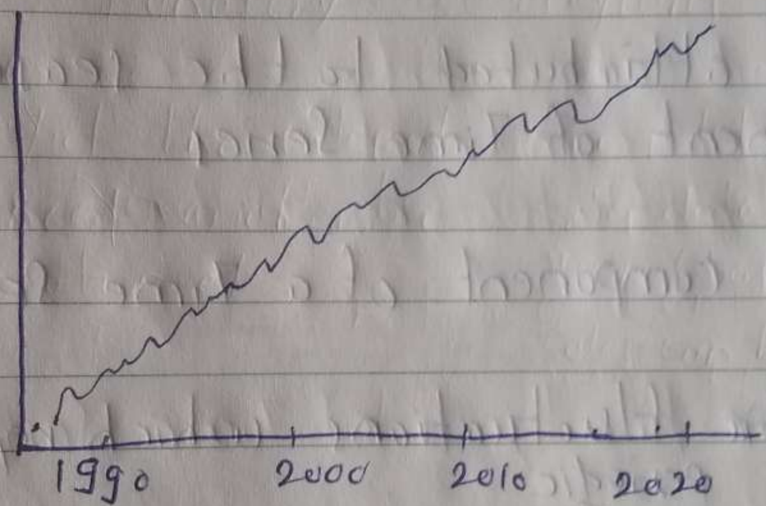
$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Components of a Time Series

1) Trend

e.g. Consumption of electricity over last 30 Years



There is a long term pattern

Reason:

Increase in sales of electronic items over a period of time

2) Seasonality

e.g. Sales of air conditioner

Sales peak in April to June year on year

This is attirbuted to the seasonal component of Time Series

Seasonal component of a Time Series -

Systematic, periodic fluctuations which repeat over time

3) Random fluctuations

fluctuations that cannot be attributed to trend or seasonal

4) Cyclic

Occurs when data exhibits fluctuations that are

- Not in a fixed time period
- Usually atleast 2 Years

- Different from seasonal behaviour

- Tend to be longer than seasonal patterns

- Exhibits more variability with respect to the seasonal patterns

Level of Time Series: Amount of fluctuations as we progress with Time

In R, when a Time Series is sliced using a decomposed function

Orignial Time Series

→ Trend componenet
→ Seasonal component
→ Random fluctuations

There are two kinds of Time Series

1) Additive Time Series

2) ~~Multiplicati~~

- mean does not change over time

- As The levels of Time Series ~~does not change~~ Progresses, fluctuations remains more or less constant

- All the components: Trend, seasonality and random fluctuations can be added up to form Observed Time Series (Additive)

$$\hat{E}_t = Y_t - \hat{T}_t - \hat{S}_t \quad (T + S + R = 0)$$

2) Multiplicatives Time Series

- As the level of time series changes, the fluctuations change

$$T * S * R = 0$$

$$Y_t = \cancel{E_t}\, S_t * T_t * E_t$$

$$\log Y_t = \log S_t + \log T_t + \log E_t$$

De-Seasonalised Series

For the Additive Model

$Y_t - S_t$ (Observed - seasonality)

For the Multiplicative Model

$\frac{Y_t}{S_t}$ (Observed / seasonality)

Steps in Time Series Analysis in R.

1) Read the data

2) Store in Time Series object

3) ~~Pass~~ save the data into object sales

Sales <- scan (http://robjhyndman.com/tsdldata/data/sales.dat)

3) Pass this object into TS function

Sales Time Series <- ts (sales)

what does that do?

helps to convert the class of data frame into time series object

why?

Functionalities within time series packages requires data to be in Time Series object

4) Pass frequency of Time Series

```
Salestime Series <- ts (sales, frequency=12)
```

Data Collected	Frequency
every month	12
every quarter	4
annually	1

5) Specify start of time period along with sub period

```
Salestime Series <- ts (sales, frequency=12,
                        start= c(1987, 1))
```

subperiod = 1 (monthly level)

6) Use plot function to plot Time Series
plot.ts(sts)

7) Decompose function to decompose various components of Time Series

decompose(salestimeSeries, type="multiplicative")

Plotting of Time Series Analysis in R

- Transform the Time Series
- Either by using a log transformation
- or by using lambda from the family of BoxCox Transformation
- For better forecast

Plot of Time Series

↓

Additive ← Transform — Multiplicative

1) Decompose
2) check presense of Trend, Seasonal, Random components
3) Select modelling technique
4) Build a model
5) Forecast
6) Validate

Forecasting Techniques

1) Exponenttial Smoothning

2) ARIMA → Does not take into account Seasonality

3) X-12 ARIMA → montly & quarterly seasonality

4) STL → Seasonal & Trend decomposition using Lewis

Simple forecasting methods

Average method	Naive method
uses average of all observations	uses most recent observation

Problems:
Information is lost when averaged or most recent observation is used

Solution:

Simple Exponential Smoothning method

- Forecasts are based on observations
- More importance given to most recent observations
- No Trend or Seasonality

- Cafe data from fpp package

Validation of Forecast

Forecast Error = Diff between observed value & Point forecast

Actual value - Predicted value

Measures to calculate forecast accuracy

1) Mean Absolute Error → MAE = Mean ($|e_i|$)

2) Root Mean Squared Error → $\sqrt{mean (e_i)^2}$

3) Percentage error → $P_i = \frac{100 e_i}{Y_i}$

All of these are scale dependent

So it is impossible to compare forecasts on time series on different scales

Where one is original Time series and

other is Log Transformed Time Series

So soln is:

4) Mean Absolute Percentage Error (MAPE)

mean (|P%|) # most commonly used

5) Mean Absolute Scaled Error (MASE)

$$MASE = \frac{MAE}{q}$$

q → scaling constant

* Ideally MAPE should be less than 7 or 7.3

* Forecast which gives least values is preferred

Residual checks

A Residual is the difference between an observation and its fitted value (which is different from forecasted value)

Forecast → Future Time point

Fitted value → Specific time points within the actual data used for model

Assumptions in Residual forecasting

1) Residuals should be uncorrelated

Test for Heteroscedasticity → Correlated or indicate a patten

→ complete info is not captured in forecast

→ Components of the forecast are missing

2) Residuals mean is close to zero

Residuals have mean other than zero → Forecasts are biased

3) Residuals have a constant variance

4) Residuals are normally distributed

* First 2 assumptions are to check the accuracy of the model

* Next ~~are~~ 2 ~~to~~ for testing or computing prediction intervals : Based on sampling distribution of the data

* Residuals that are normally distributed indicate that they are Identically or Independently Distributed Data (I.I.D's)