Introduction to Statistical Methods

Moving Average (MA) & Simple Exponential Smoothing (SES)

Revision-2.0

Prof Vineet Garg

BITS Pilani Work Integrated Learning Programmes (WILP)
Bangalore Professional Development Center

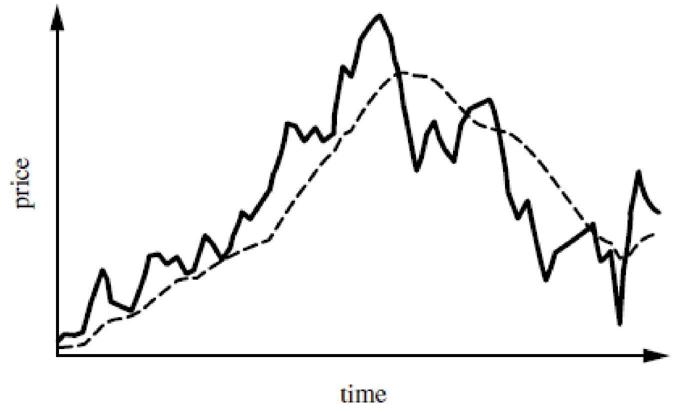
Introduction

- In Statistics and Data Science, there is an important concept of *Temporal Data Sets* that has two types of data:
  - Time Series Data: it is normally a sequence of real numbers that vary with time.
  - Sequence Data: it is a sequence of ordered events with or without the concrete notion of time. For example: web page traversal sequence.
- A time series data consists of sequence of values obtained over *repeated measurement of time*.
- The examples where time series data is generated:
  - Demand forecast
  - Stock market analysis
  - Yield projections
  - Inventory studies
  - Natural phenomena (e.g. seismic activities)

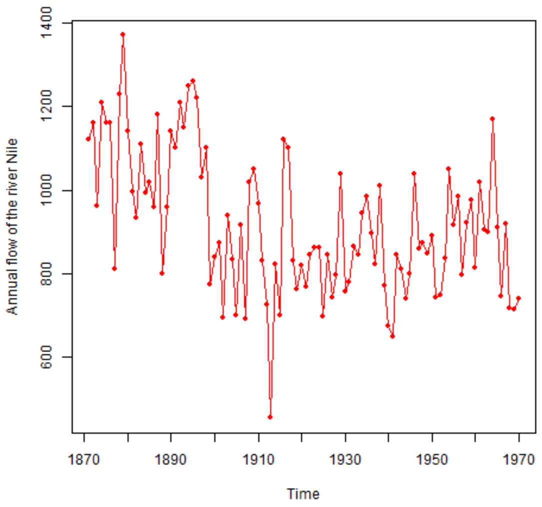
- There are two primary goals in the time series analysis:
  - Modelling: to gain insight into the underlying forces that generate the time series.
  - Forecasting: predict the future values of the time series variables.

Trend Analysis with Moving Average

- Trend analysis in time series data indicate the *general direction* in which the time series graph is moving over a long interval of time.
- For example in the shown graph the dashed line shows the general trend.
- Finding out the general trend is harder than saying.



- For example, the [Nile river data]() set contains the data of the Nile river annual flow under the Aswan dam in Egypt over a period of 1871–1970 in the units of 10^8 cubic meters.
- How can we draw the general trend of the data like this?
- Is it a work of fine arts or are there mathematical techniques to identify the general trend?



Moving Average Smoothing

- Suppose that there are n time periods denoted by $t_1, t_2, t_3, \ldots, t_n$ and the corresponding values of the Y variable are $Y_1, Y_2, Y_3, \ldots, Y_n$.
- Decide on the period (k) on moving averages. For the short time series k = 3 or 4 may be sufficient. For long time series, k = 12, 15 etc. may be required.
- k-Moving Average is calculated as shown in the table below (notice the difference when k is odd or k is even):

Years (t)	Variable (Y)	k=3 Moving Average	k=4 Moving Average	k=4 Moving Average Centered
t_1	Y_1	.	.	.
t_2	Y_2	$Y_1+Y_2+Y_3$ / 3	$X_1 = Y_1+Y_2+Y_3+Y_4$ / 4	.
t_3	Y_3	$Y_2+Y_3+Y_4$ / 3	$X_2 = Y_2+Y_3+Y_4+Y_5$ / 4	X_1+X_2 / 2
t_4	Y_4	$Y_3+Y_4+Y_5$ / 3	$X_3 = Y_3+Y_4+Y_5+Y_6$ / 4	X_2+X_3 / 2
.	.	.	.	.
.	.	.	.	.
t_{n-2}	Y_{n-2}	.	.	.
t_{n-1}	Y_{n-1}	$Y_{n-2}+Y_{n-1}+Y_n$ / 3	.	.
t_n	Y_n	.	.	.

Details of Calculations

Moving Average Smoothing

The following examples explain how moving average is calculated for odd and even value of k.

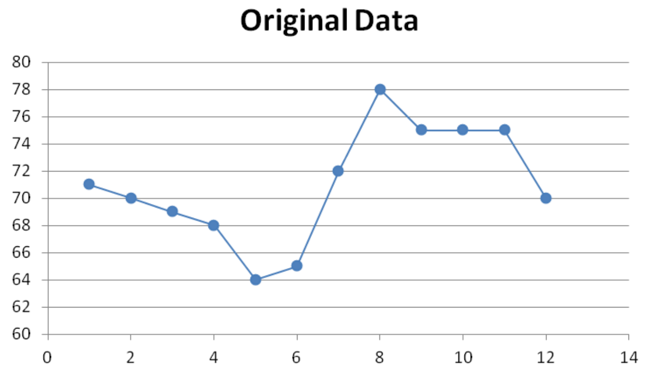
Time Unit	Variable	k=3 Moving Average
1	9	
2	8	8.67
3	9	9.67
4	12	10.00
5	9	11.00
6	12	10.67
7	11	

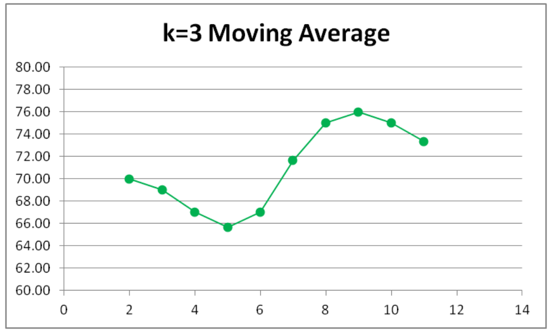
Time Unit	Variable	k=4 Moving Average	k=4 Moving Average Centred
1	9		
1.5			
2	8		
2.5		9.5	
3	9		9.5
3.5		9.5	
4	12		10.00
4.5		10.50	
5	9		10.75
5.5		11.00	
6	12		
6.5			
7	11		

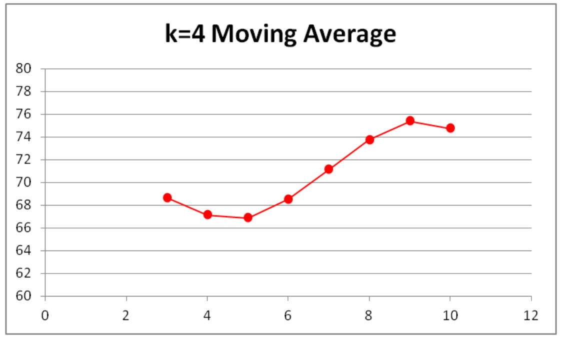
Illustration

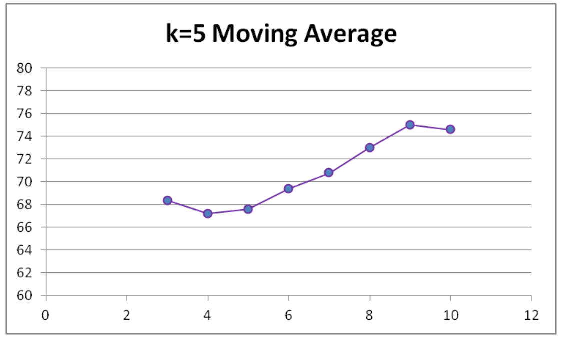
Moving Average

#	Variable	k=3	k=4		k=5
1	71				
2	70	70.00	69.50		
3	69	69.00	67.75	68.63	68.4
4	68	67.00	66.50	67.13	67.2
5	64	65.67	67.25	66.88	67.6
6	65	67.00	69.75	68.5	69.4
7	72	71.67	72.50	71.13	70.8
8	78	75.00	75.00	73.75	73
9	75	76.00	75.75	75.38	75
10	75	75.00	73.75	74.75	74.6
11	75	73.33			
12	70				

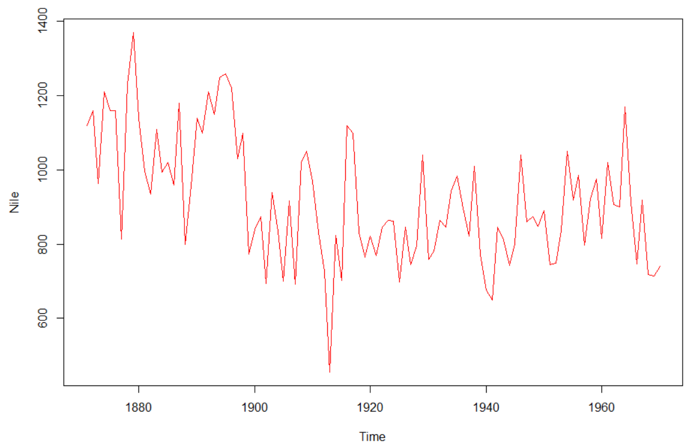




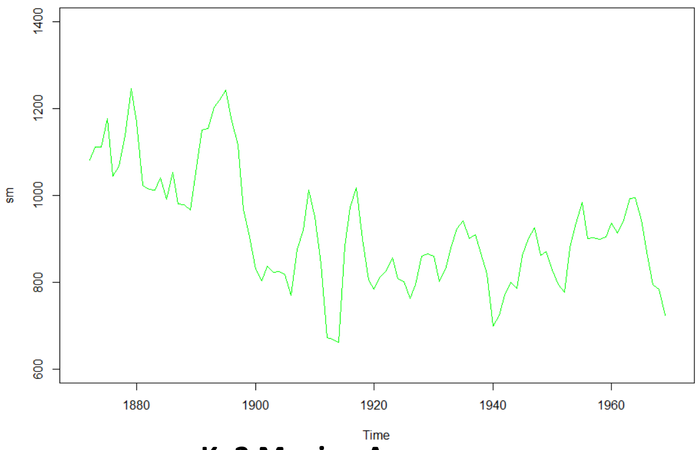




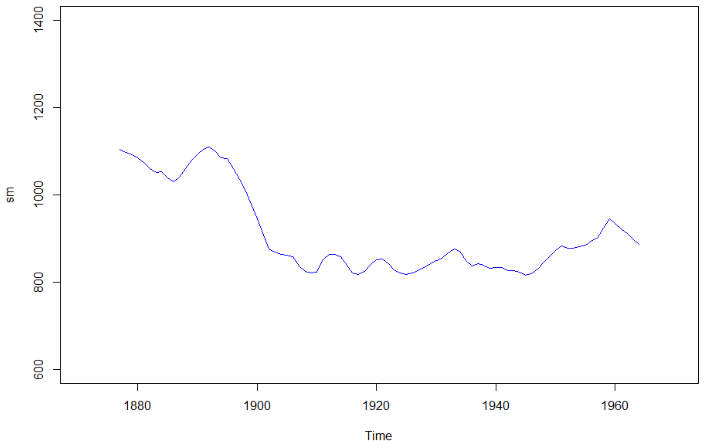
Example: Nile Dataset - Moving Average



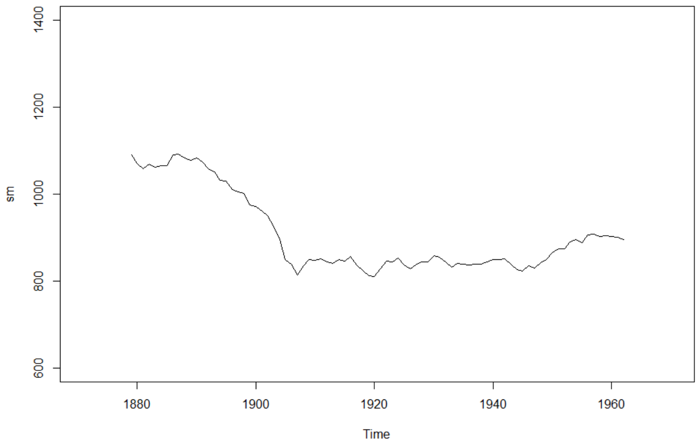
Original Data



K=3 Moving Average



K=12 Moving Average



K=17 Moving Average

Forecasting Using Moving Average

Forecasting (F) using the moving average and actual values (D) can be done using the following formula for the interval size of N for calculating the average:

$$F_{t+1} = \frac{1}{N} \sum_{k=t+1-N}^{t} D_k$$

- Let us say forecast F_8 is to be calculated using the previous 5 actual values. So t+1 = 8 and N = 5.
- The actual values that will be used to calculate the moving average: (t+1-N) = 8-5 = 3 to t = 7 or D_3, D_4, D_5, D_6 and D_7.

Limitations

Moving Average

- Identifies only the superficial trend.
- Suppresses majority of minor peaks and dips, only shows the major ones.
- Loses few data points.
- Mostly left to the user’s interpretation.
- No mathematical measure for forecasting - *speculative*.

Simple Exponential Smoothing

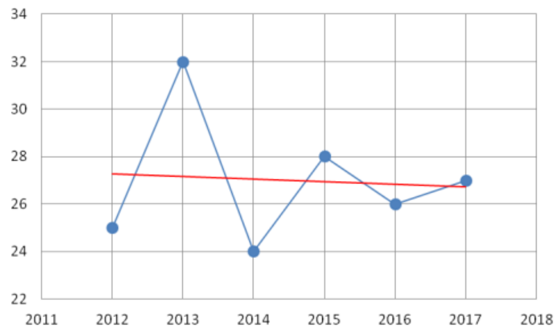
Problem Statement

- The demand data for six years is provided from the year 2012 to 2017 in the given table.
- The forecast for the year 2018 is to be found out.
- How would you go about forecasting it?
    - Average of the 6 years demands = 27 ?
    - Average only of the last 4 years demands considering them to be more relevant = 26.25?
    - Considering last two years demands increasing trend = 28?
    - Considering the increase/decrease alternate trend = 26?
    - Weighted average of the last 3 years demands - weight for 2015 is 1, 2016 is 2 and 2017 as 3 = 26.83?
    - Eliminating 2013 demand as an outlier and considering the average of others = 26?
    - Another value X, considering other factors e.g. increased purchasing power etc.?

Year	Demand
2012	25
2013	32
2014	24
2015	28
2016	26
2017	27
2018	?

Simple Exponential Smoothing

Explanation



- In the previous slide, forecast is based upon:
  - Simple arithmetic average
  - Selective average (or moving average) for few recent years
  - Weighted average
  - Considering some kind of trend
  - Other methods – considering external factors, removing outliers etc.
- Is there any *distinguishable trend* in the provided data that cannot be attributed to randomness?
- The answer is probably no. Such models are called level models or constant models. Where forecast (F) can be written as F = a + ε, where a is some constant and ε is the noise that is assumed to get cancelled over a large sample of data points with some variance, i.e. its average will be 0 and variance will be σ^2.
- ***Fluctuations around a slowly-varying mean.***
- When only the last few years are considered skipping the first *k* years to calculate the average, the situation can be called that it is a *k moving simple average*.
- This situation is also equivalent to that - 0 weights are assigned to the older *k* years and weights 1 are assigned to the recent years - *a form of weighted average.*
- The combination of moving and weighted average was also exhibited in the last slide when the weighted average of the last 3 years was calculated.
- There is a drawback in this approach – it simply assign 0 weights to the older data.
- Can there be a model where all the data points are considered and progressively increasing weights can be provided to more recent data?
- The answer is yes and such models are called ***Exponential Smoothing Models.***

Simple Exponential Smoothing

Modelling

- Exponential smoothing models can be written as:

$$F_{t+1} = \alpha.D_t + (1- \alpha).F_t$$

Where, F is the forecast, D is the demand, t is the time and α is the smoothing constant with $0 < \alpha < 1$.

- In the given data the F7 needs to be found out. The smoothing equations can be written as:

  $F_7 = \alpha.D_6 + (1- \alpha).F_6$

  $F_6 = \alpha.D_5 + (1- \alpha).F_5$

  ……….

  $F_2 = \alpha.D_1 + (1- \alpha).F_1$

- If α is taken as 0.2 and F_1 is taken as 27 (the average demand for 6 years), the forecast for each year (and for the 2018 too!) can be calculated.

Year	Demand (D_t)	Forecast (F_t)
2012	25	27.00
2013	32	26.60
2014	24	27.68
2015	28	26.94
2016	26	27.16
2017	27	26.92
2018		**26.94**

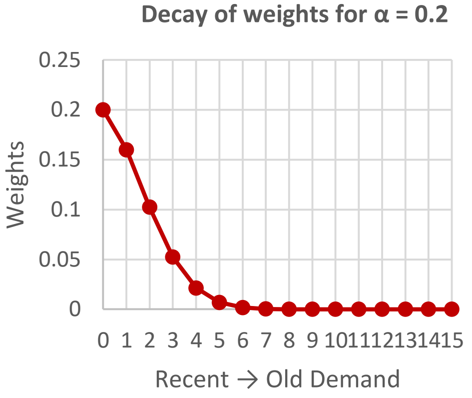
- Advantages:
  - ✓ All the data points are used.
  - ✓ Progressively increasing weights are assigned to more recent data.

Simple Exponential Smoothing

Working Details

$F_7 = \alpha.D_6 + (1-\alpha).F_6$
$= \alpha.D_6 + (1-\alpha).[\alpha.D_5 + (1-\alpha).F_5]$
$= \alpha.D_6 + \alpha.(1-\alpha).D_5 + (1-\alpha)^2.F_5$
$= \alpha.D_6 + \alpha.(1-\alpha).D_5 + (1-\alpha)^2.[\alpha.D_4 + (1-\alpha).F_4]$
$= \alpha.D_6 + \alpha.(1-\alpha).D_5 + \alpha.(1-\alpha)^2.D_4 + (1-\alpha)^3.F_4$
\- - - - -
$= \alpha.D_6 + \alpha.(1-\alpha).D_5 + \alpha.(1-\alpha)^2.D_4 + ... + \alpha.(1-\alpha)^5.D_1 + (1-\alpha)^6.F_1$

If $\alpha = 0.2$:
$\alpha.(1-\alpha) = 0.16$,
$\alpha.(1-\alpha)^2 = 0.13$
$\alpha.(1-\alpha)^3 = 0.10$
……..



Since $0 < \alpha < 1$, observe that $\alpha > \alpha.(1-\alpha) > \alpha.(1-\alpha)^2 > .... > \alpha.(1-\alpha)^5$

- ✓ Hence it is progressively increasing weight to the more recent data.

Now, for the weighted average, the weighted sum needs to be divided by the sum of weights,
The sum of weights assigned to the demands (leaving out the last term) =
$\alpha + \alpha.(1-\alpha) + \alpha.(1-\alpha)^2 + ... + \alpha.(1-\alpha)^5$
For, large set of data sum would be equal to $= \alpha / 1-(1-\alpha) = 1$ (sum of geometric progression of infinite terms)

- ✓ Hence there is no need to divide by the sum of weights because the sum of weights is 1.

Why α is chosen between 0 and 1 in $F_{t+1} = \alpha.D_t + (1-\alpha).F_t$

- ✓ If $\alpha > 1$, the second term would have the negative impact
- ✓ if $\alpha < 0$ then the first term would.

For the large dataset the last term $(1-\alpha)^6.F_1$ will have the minimum impact because $(1-\alpha)^n$ will be nearly zero, where n is a very large. So, the accuracy in the initial forecast (F_1) is not *that* desirable.

- ✓ However F_1 should be derived from some logical method like average of data points or as the last available data point etc.

Exercise

Table shows the crude oil production of an OPEC member for seven consecutive weeks. Forecast the oil production for the 8th week using simple exponential smoothing procedure. Assume the values of the relevant parameters (if you need any) and ignore any external factor or dependency that might impact the production.

Week		1	2	3	4	5	6	7
Crude Oil Production	*(million tonnes)*	40	42	39	41	38	42	39

Answer: Considering forecast for the 1st week as the average demand for the first 7 weeks and smoothing constant as 0.2, the forecast for the 8th week = **40.05.**

References

1. NIST Handbook – Introduction to Time Series Analysis
2. NPTEL Videos – Forecasting by Prof G. Srinivasan

Appendix

(Other Prediction and Forecasting Techniques)

For Self Study

Will Not be Included in the Evaluation Components

Holt's Model

- Holt's model extends the concept of simple exponential smoothing to the dataset which is having a trend.
- It considers all the data points and assigns progressively more weightage to the recent data points and lesser weightage to the older data points.
- According to the Holt's model, the forecast at point (t+1) is given by:

  $F_{t+1} = a_t + b_t$

  where a_t is the level of the smoothed value up-to and including the last data point and b_t is the slope of the line at point t.
- The values of a_t and b_t are updated using the following equations and smoothing constants (α and β each of which is between 0 and 1):

  $a_t = \alpha.D_t + (1- \alpha).F_t$ --------------------------------------(i)

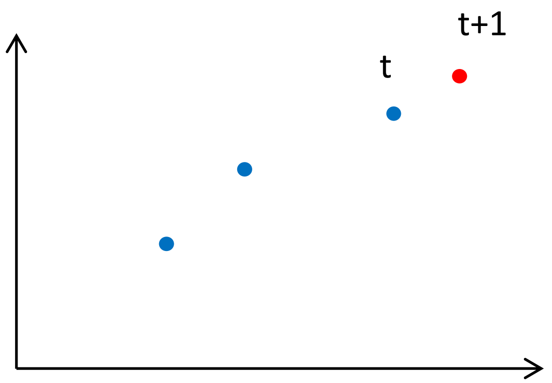
  $b_t = \beta.(a_t - a_{t-1}) + (1 - \beta).b_{t-1}$ ----------------------------(ii)
- In the first equation to calculate the level component (a_t), α weightage is assigned to the demand and (1- α) weightage is assigned to the forecast at time t.
- In the second equation, to calculate the slope (b_t), β weightage is assigned to the difference of the level component and (1- β) weightage is assigned to the last slope.
- In the Holt's model an attempt is made to find out the level and slope at each point in order to forecast.

Holt's Model

Explanation

- Demand and Forecast for up-to the point t is known and the forecast for the point (t+1) is to be found out.
- From, Y = a+bt, the forecast for t+1 can be written as:

  $F_{t+1} = a_t + b_t.1$



- While linear regression assumes the same slope all through out, Holt's model adapt the *a* and *b* components through double exponential smoothing as:

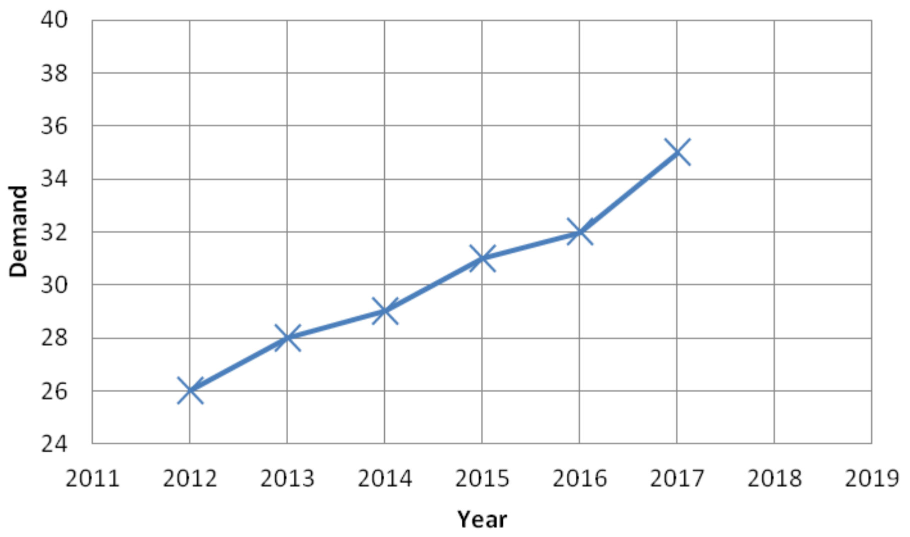
$a_t = \alpha.D_t + (1- \alpha).F_t$ --(i)

$b_t = \beta.(a_t - a_{t-1}) + (1 - \beta).b_{t-1}$ ----------------------------(ii)

Illustration

Holt's Model

Year	Demand (D_t)
2012	26
2013	28
2014	29
2015	31
2016	32
2017	35
2018	?



To start with a_1 is taken as D_1 that is = 26 and b_1 is taken as the slope $(D_6\text{-}D_1)/5 = (35\text{-}26)/5 = 1.80$, $\alpha = 0.2$ and $\beta = 0.3$. Using these a_1 and b_1, subsequent values of a, b and F can be calculated using the equations:

$F_{t+1} = a_t + b_t$

$a_t = \alpha.D_t + (1\text{-}\ \alpha).F_t$

$b_t = \beta.(a_t - a_{t-1}) + (1 - \beta).b_{t-1}$

t	D_t	F_t	a_t	b_t
1	26	*	26.00	1.80
2	28	27.80	27.84	1.81
3	29	29.65	29.52	1.77
4	31	31.29	31.24	1.76
5	32	32.99	32.79	1.70
6	35	34.49	34.59	1.73
7		36.32		

Seasonality

- Data is provided for 3 years divided into 4 quarters. Quarterly forecast for the 4th year is to be found out.
- An observation indicates that for each quarter there is an increasing trend year-over-year and that trend is reflecting in the total also.
- However, for any year the quarters are not showing any trend or constant levels. There is a variation.
- This behaviour in the data is termed as *seasonality*. This also demands forecasting models that captures seasonality. Sale of several products exhibit seasonality like winter/summer products.
- For a simple seasonality model, a seasonality index (SI) is calculated for each year and each quarter. For example:
  - Y-1, Q-1, the SI = 53/157 = 0.34
  - Y-3, Q-4, the SI = 56/189 = 0.30
  - so on....
- Average SI is also calculated for each quarter. It can be said that Q1 accounts for 33.4% of the sales, Q2 accounts for 14.2% and so on.
- The sums of each quarter for the years show an increasing trend (157, 173, 189). Which can be used with linear regression or Holt's model to forecast for the next year.
- Using an appropriate model, when we get the forecast for the next year, average SI can be used to decompose to the quarterly projection.

	Year-1	Year-2	Year-3
Q-1	53	58	62
Q-2	22	25	27
Q-3	37	40	44
Q-4	45	50	56
∑	**157**	**173**	**189**

Seasonality Index (SI)				
	Year-1	Year-2	Year-3	Average
Q-1	0.34	0.34	0.33	0.334
Q-2	0.14	0.14	0.14	0.142
Q-3	0.24	0.23	0.23	0.233
Q-4	0.29	0.29	0.30	0.291

Exercise

1. For the data provided in the simple seasonality model on the last slide, forecast the quarterly projection of the fourth year using linear regression.

   (**Answer**: equation of the best fitting line: Y = 141 + 16.t
   Therefore, yearly forecast for the 4^{th} year = 141+16x4 = 205
   Quarterly projection for the 4^{th} year: Q1= 68.39, Q2= 29.21, Q3 = 47.81, Q3 = 59.58 using the average SI)

2. In the simple seasonality model data, each quarter is showing the increasing trend year-over-year. Why it would be inappropriate to use the linear regression for each quarter independently and forecast the quarterly values for the 4^{th} year?

   (**Answer**: It is possible mathematically, but dependency among the quarters will not be captured. Secondly, if we consider Q1 data for three years to project the Q1 value for the 4th year it would be equivalent of saying that only 1st,5th and 9th data points are considered out of total 12 data points to project for the 13th data point and that would be a lossy forecast.)

Winter's Model

- Following the similar lines of simple exponential and Holt's model, when there is seasonality in the data, there are three factors that need to be progressively adapted: level, slope and seasonality.
- This phenomenon is captured by Winter's model. The equations of Winter's model is given by:

  $a_{t+1} = \alpha.(D_{t+1}/c_{t+1}) + (1- \alpha).(a_t + b_t)$ *for level*

  $b_{t+1} = \beta.(a_{t+1}- a_t) + (1- \beta).b_t$ *for slope*

  $c_{t+p+1} = \gamma.(D_{t+1}/a_{t+1}) + (1- \gamma).c_{t+1}$ *for seasonality*

  $F_{t+1} = (a_t + b_t).c_{t+1}$ *forecast*

  Where α, β and γ are smoothing constants between 0 and 1 and p is the period of seasons. For example p = 4 for four quarters.
- Also note that c, D and F are seasonal (quarterly for the shown example) components while all others are yearly components.

Illustration

Winter's Model

	Year-1	Year-2	Year-3
Q-1	53	58	62
Q-2	22	25	27
Q-3	37	40	44
Q-4	45	50	56
∑	157	173	189

Seasonality Index (SI)			
	Year-1	Year-2	Year-3
Q-1	0.34	0.34	0.33
Q-2	0.14	0.14	0.14
Q-3	0.24	0.23	0.23
Q-4	0.29	0.29	0.30

$a_{t+1} = \alpha.(D_{t+1}/c_{t+1}) + (1-\alpha).(a_t + b_t)$

$b_{t+1} = \beta.(a_{t+1} - a_t) + (1-\beta).b_t$

$c_{t+p+1} = \gamma.(D_{t+1}/a_{t+1}) + (1-\gamma).c_{t+1}$

$F_{t+1} = (a_t + b_t).c_{t+1}$

Initialization:

$\alpha = 0.2$, $\beta = 0.3$ and $\gamma = 0.25$

For seasonality: $c_1=0.34$, $c_2=0.14$, $c_3=0.24$, $c_4=0.29$, $c_5= 0.34$

For level: $a_1 = D_1/c_1 = 53/0.34 = 156$

For slope: $b_1 = (173\text{-}157=16)\ /\ 4$ periods $= 4$

Example Calculation:

$a_2 = 0.2 \times (22/0.14) + 0.8 \times (156+4) = 159.43$

$b_2 = 0.3 \times (159.43 - 156) + 0.7 \times 4 = 3.83$

$c_6 = 0.25 \times (22/159.43) + 0.75 \times 0.14 = 0.1395$

$F_2 = (156+4) \times 0.14 = 22.40$

$F_3 = (159.43+3.83) \times 0.24 = 39.18$ *other forecast values can be similarly calculated.*

Thank You