5) STA 2401
TIME SERIES

April ... we... → type regress analysis

THE TIME SERIES

A T.S is a set of numerical values or observations of a group variable listed at ~~successive~~ sequentially intervals of time which means that the data regarding the variables is listed in chronological order.

X_t	X_{t+1}
regressing a lack generating	from original

* it is contineous

From the definition an important feature of T.S is that the observations are correlated.

Types of time series – data dealing with economy

1) Economic time series – Many T.S arise in economics e.g share prices on succesive ~~dealings~~ days, export totals in succesive months, monthly inflations, yearly G.D.P, balance of trade, tax rates. share prices on successive days, export totals " " months, average incomes in succe mon...

2) Physical time series – Many T.S occur in physical ~~services~~ sciences particularly in marine services and meteorology. examples rainfall on succesive days; temperature in successive hours etc.

3) Marketting Time Series – Examples; sales time series, advertising *Commerce features of population. – sales you get – analysis of sales figures in succesive weeks

4) Demographic Time Series – time series occurs in the study of populations e.g. population of country in 10 years.

Objectives of time series analysis (TSA) trend – long term change in the mean level.

There a nº of t.s.a.

This are grouped in three 1) description
2) prediction
3) explanation
D E P

1.) Description
When presented with T.S,

-> what type of trend (increasing, decreasing)

-> what seasonal effect.

~~how to ident~~

First step is to plot the data and obtain simple descriptive measures of the main properties of the series.

- Features such as the trend and seasonal effects can sometimes be seen from the plot.

-> The plot/graph normally shows trend and seasonal effect but sometimes some unusual observations are normally seen. (outliers)

plot
obtain data.

outlier

2) Explanation.

- When observations are taken on 2 or more variables, it may be possible to use the variation in 1 variable to explain the variation to another variable.

- This may lead to understanding the mechanism that generated the series

3) Predictions e.g. sales
predict based on observed data

Giving an observed T.S, one may want to predict the future values of the series.

Approaches to T.S.A.

Autocorrelation - Time domain
Spectral density - Frequency domain

- There are 2 way approaches to T.S.A.

- In the first approach, main tool of analysis is called Autocorrelation function (ACF)

- The ACF describes the evolution of a process through time.

Inference based on ACF is called Time Series Analysis in the (time domain.)

- The ~~othe~~ other approach depends on the Spectral Density function which describes how variation in a series may be explained by variations at different frequencies.

2 - spectral analysis

- Analysis based on S.D.F is called (the analysis in frequency domain)

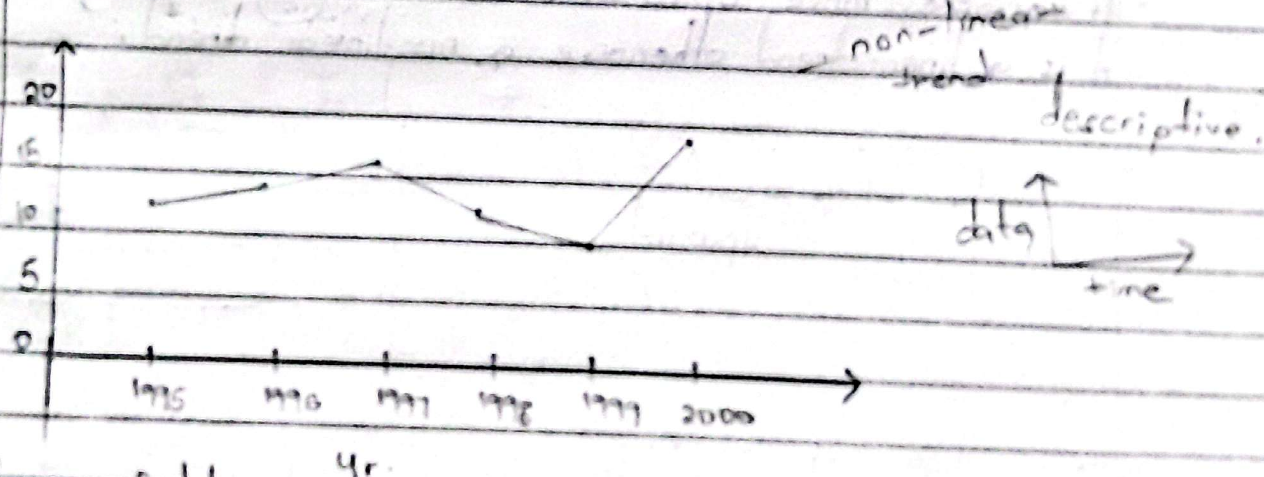
- Given a time series we need to study about the forces that influence the variations in the series over the given period of time.
- Consider the sales of T.V in (thousands) by a producing company.

Table 1

Yr	1995	1996	1997	1998	1999	2000
Nº of TV sets sold in ('000)	12	14	16	12	10	18

@ 1995-2000.

- Analyze the data and give some trends regarding the sales.
- Also the company would want to know why sales dropped in 1998, 1999 and why did it increase afterwards.
- In other words, company would like to explain forces that affect the sales.
- The following are the 4 changes in the values of the variables.



tendencies of data

1) Changes which occured due to general tendencies of the data to increase or decrease.

2) Changes which occured due to change in climate, weather conditions and festivals.

3) Changes which occured due to booms and depressions.

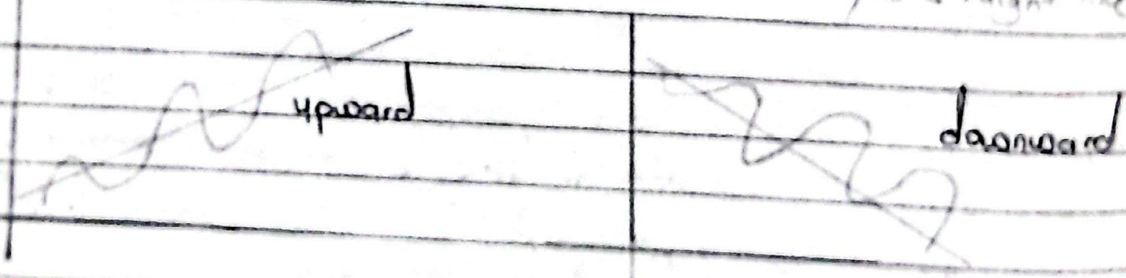
4) Changes which occured due to some unpredictable forces

Components of time series. Types of variation.

- The behaviour of a T.S over periods of time is called the movement of T.S
- T.S is classified into the following 4 components. behaviour of T.S over a period of time.
  1) Long-term trend / Secular trend.
  2) Seasonal variations
  3) Cyclic variations
  4) Irregular / Random variations.

1) Long term / Secular trend.

- This refers to the upward / downwards trends of the series.
- The movement can be characterized by a trend curve.
- If the curve is a straight line, it is called a trend line.
- If the variables increases over a long period of time, then it is an upward trend.
- Variables decreases over a long period - donward trend
- If variables move upwards donwards along a straight line it is a linear trend otherwise a non-linear trend.
  ↳ not straight line

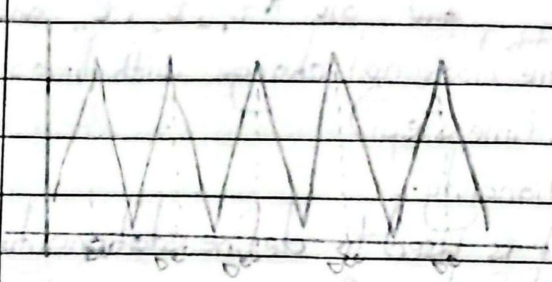


2) Seasonal variations.

- Variation in T.S that are periodic in nature and occur regularly over a short period of time during an year.
- Variations can be precise and can be forcasted
- i.e., Prices of vegetables drop after a rainy season and up during summer every year.

* stationary data -

; Prices of cooking oil reduces after harvest of oil seeds and goes up after sometime.



3) Cyclic variations. → e.g economic data are sometimes thought to be affected by biz cycles; period btwn 5-7 yrs.

- Long term variations that represent consistent rise and decline in the values of the variables.
- Since this are long-term variations in the T.S, period of variation is usually > one year.

4) Random variations: - unpredictable in nature.

- They are called irregular movements.
- Mvmt that occur usually in brief period of time without any pattern and that are unpredictable in nature.
- This movement will not have any regular period time of occurence; e.g. national strikes, floods, earthquakes, outbreak of deaseses (COVID), wars.
- It is difficult to study the behaviour of such (time) series.

Stationarity. - if statistical values of the time series are constant in time. independent of time. { constant mean, " variance, acf.

- A T.S is stationary if it has no systematic change in mean (no trend), no systematic change in variance and if they are no periodic variations.

- no change in mean
- " variance
- no periodic variation

- Mathematical definition on stationarity.

* turning a non-stationary series to a stationary one.

$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$ diff b/wn 1st & 2nd order stationarity

1) Strictly stationary / First order stationarity.

replace y Definition 1: A T.S is said to be strictly stationary if the joint distribution of $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ is the same as $x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_n+h}$ for all t_1, t_2, t_n and h. That is the probability structure does not change with time.

- h is refered to as the lag (time step).

Weakly stationary / Second order stationarity.

Definition 2: In practice it is used to define stationarity in less stringent way than given in definition 1.

- A T.S is weakly stationary is mean is constant and autocovariance function (ACVF) is independent of time but depend only on the distances b/wn the variables.

Remark → definition 1 is defined as 1st order stationarity
" 2 " " " 2nd "

→ 2nd order is more practical as compared to the 1st.
→ In case of 2nd order stationarity the ACVF is denoted and defined as

$\gamma(h) = Cov(X_t, X_{t+h})$ assumption of normality

$= E(X_t, X_{t+h}) - E(X_t)E(X_{t+h})$ ---- (1.1) mean (0)

- Since the mean of a stationary process is constant, it can be taken as zero hence we have

$\gamma(h) = E(X_t, X_{t+h}) - 0$ ---- (1.2)

- We note that is influenced by the units of measurement. To compare the basic properties T.S, it is important to have a fn that is not influenced by units of measurement.

- Such a fn is a theoritical autocorrelation function which is defined as no element of time → independent on time → not influenced by units of measurement

$\rho_h = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{Var(X_t)}$
acf

- Clearly the acf is independent of time.
- The ACF is important for describing the T.S

n graphs - clear title
State units of measurements T.S. economic data.
clearly label the axes

- It is also an important tool for assecing the properties of the T.S model.

e; Henceforth we shall take stationarity to mean that stationarity in the weak state.

long-term trend / secular trend one of 4 components.

Methods ~~used~~ of fitting/measuring trend.

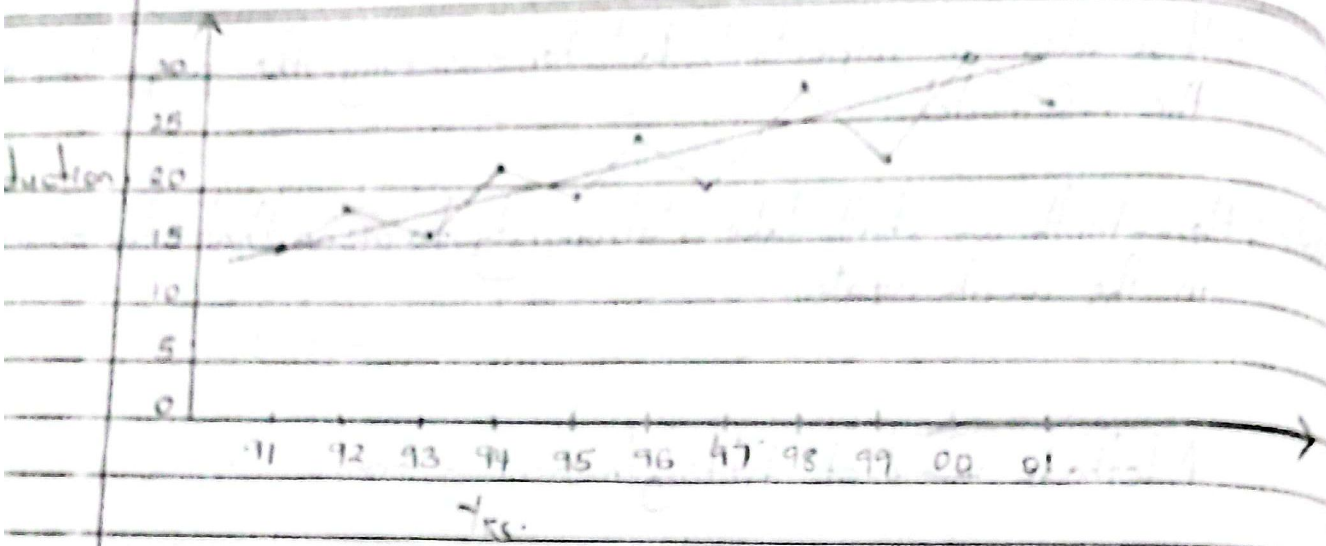
1. Freehand/Graphic methods. F
2. Semi average methods F
3. Moving averages methods M
4. Method of least squares M
5. Filtering S

1) Freehand/Graphic methods.

- This is simplest method of drawing a trend curve.
- We plot the values of the variable against time on the graph and join the points.
- The trendline is then fitted by inspecting the graph of the time series.
- Fitting the trend line by this method is arbitrary.
- Trend line is drawn such that the n° of fluctuations on either sides are approximately the same.
- The trend line should be a smooth curve.

Example 1
Find the trend with the help of a free hand curve method for the data below.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
	15	18	16	22	19	24	20	28	22	30	26



b. Semi average methods

The methods of fitting a linear trend with the help of semi average methods are as follows

a.) When n° of yrs is even.

- When n° of yrs is even, the T.S. is divided into two equal parts.
- The total in each part is estimated and divided by the n° of items to obtain the arithmetic means of the two parts.
- Each average ~~it~~ is centred in the period of time it has been computed and ploted on the graph.
- A straight line is drawn passing through this points.
- This is the required trend line.

b) When n° of yrs is odd.

- When n° of yrs is odd, the value of the middle yr is ommited to divide the T.S in two equal parts.

The ~~precid~~ procedure above is followed

The figures given below shows the export of sugar from India from 1971 to 1980. Determine the trend by using semi average method.

4.02 * 5.4

1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
3.9	1.3	1.1	4.4	9.4	9.6	3.4	2.5	8.6	2.9

The n° of yrs is 10, that is it is even.
The series is divided into two parts consisting of first and last 5 yrs.

Average for first 5 yrs →

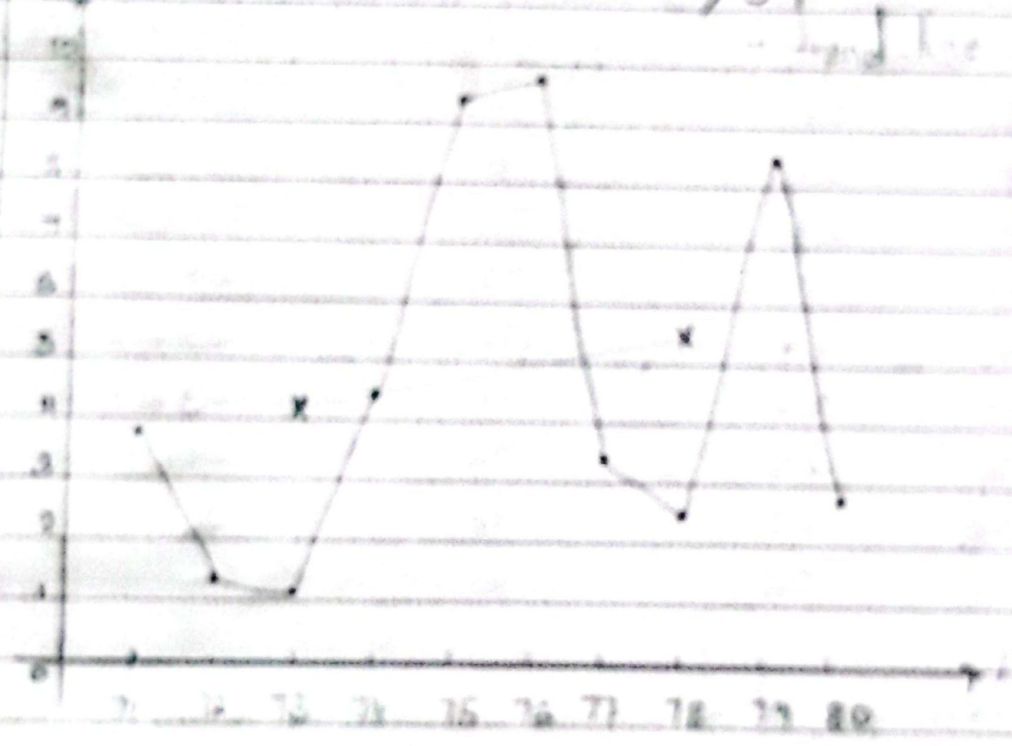
$\frac{1}{5}(3.9+1.3+1.1+4.4+9.4)$

$\Rightarrow 4.02$

Average for last 5 yrs

$\frac{1}{5}(9.6+3.4+2.5+8.6+2.9)$

$\Rightarrow 5.4$



3) Moving averages method.

a) When the period of moving average is odd

- Obtain the T.S
- Select a period of moving average such as 3, 5, 7, ...
- Compute moving totals according to the length of period of moving average.

If the length of the period of moving average is 3, it is 3 yrly moving average and moving totals of is computed as follows a+b+c, b+c+d, c+d+e, d+e+f, ...

- Similarly for a five yr moving totals
  a+b+c+d+e, b+c+d+e+f, c+d+e+f+g

Step 4: Place the moving totals at the centre of the time span from which they are computed.

Step 5: Compute moving averages by dividing the moving totals by the period of moving average.
This moving averages are called the trend values.

Example 3

Calculate the 3 yr moving averages for the data in the table below.

Yr	(Y) Production	3 yr moving totals / Moving totals	3 yr M.A (Y_c) Moving averages	$(Y - Y_c)$ Short term fluctuation
1988	21			—
1989	22	66	22	0
1990	23	70	23.33	-0.33
1991	25	72	24	1
1992	24	71	23.67	0.33 1.67
1993	22	71	23.67	-0.67
1994	25	74	24.67	0.33 1.67
1995	27	78	26	-1
1996	26	-	-	-
n=9				

b) When period of moving average is even.

- When period of moving av. is even such as 4, compute the moving averages using the following steps.

1: Obtain the T.S data
2: Obtain the length of the period of moving av. e.g. 6, 8, 10, ...
3: Compute 4 yrly moving totals and place them at the centre of the time span.

- The 4 yr moving totals are computed as follows
  a+b+c+d, b+c+d+e, c+d+e+f, ...

Compute 4 yrly averages and place them at the centre of the time span

Step 5: Take 2 period moving average of moving averages and place them at the centre of the periods.
This is called centering of moving averages.

Example The following table gives the average monthly productions (thousands) of new passenger cars btwn 1976 and 1985. Calculate 4 yrly moving averages:

(y-$\hat{y}$)	Y.	Production (y)	M.T	(Ma) M.A	(y_c) 2-item MA
	1976	~~708~~ 708			
	1977	767	2941	735.25	
50.625	1978	764	2766	691.5	713.375
41.25	1979	7082	2520	630	660.75
-54.125	1980	533	2177	544.25	587.125
-5.75	1981	521	2037	509.25	526.75
-101	1982	421	2139	534.75	522
9	1983	562	2285	571.25	553
	1984	635			
	1985	667			

n = 10

4) Method of least squares:

- In this method, the trend curve is determined by fitting a mathematical equation.
- This method is more accurate & precise and can be used in forecasting.
- Given a T.S data set, a straight line / parabolic curve can be fitted.
- There are 2 methods
  - a) Direct method
  - b) Short-cut method.

a) Direct method.

1; Convert yrs into accurate natural numbers 1, 2, 3, ... and call it x and find Σx

(2.1) --- $y = \beta_0 + \beta_1 x$ · ① → no specific value for x hence step 1 ② Sum whole equation

getting normal equation ★ — Apply summation.

$\Sigma y = n\beta_0 + \beta_1 \Sigma x$ --- ② (2.2)

$\Sigma xy = \beta_0 \Sigma x + \beta_1 \Sigma x^2$ --- ③ (2.3)

$x \quad x^2 \quad xy \quad \hat{y} \quad y = \beta_0 + \beta_1 x$

Step 2; Find the squares of x values and obtain the sum (Σx^2)

Step 3; Add the values of Y ~~and~~ to abtain sum of Y.

Step 4; Multiply x values with corresponding Y values and have the totals (Σxy)

Step 5; Substitute the values into the normal equations and solve the parameters.

Step 6; Substitute the values into the least square equation and find the trend values for the various values of x.

Y x Σx^2 Σy xY

b) Short-cut method.

- The calculations are simplified such that the midpoint in time is taken as the origin so that (sum of x) $\Sigma x = 0$.
- When $\Sigma x = 0$, the normal equations reduces to

$$\Sigma y = n\beta_0 \text{ ---- } (2.4)$$

Short-cut, make $\Sigma x = 0$.

$$\beta_0 = \frac{\Sigma y}{n}$$

$$\frac{\Sigma xy}{\Sigma x^2} = \frac{\beta_1 \Sigma x^2}{\Sigma x^2}$$

$$\beta_1 = \frac{\Sigma xy}{\Sigma x^2} \text{ --- } (2.5)$$

Example: Find the trendline by the method of least squares given the following data. (Y)

Yr	Production	X = Yr − 1979	X^2	XY
1976	672	−3	9	−2016
1977	824	−2	4	−1648
1978	967	−1	1	−967
1979	1204	0	0	0
1980	1464	1	1	1464
1981	1758	2	4	3516
1982	2057	3	9	6171
	8946	$\Sigma x = 0$	28	6520

$$\beta_0 = \frac{8946}{7} \Rightarrow 1278$$

$$\beta_1 = \frac{6520}{28} \Rightarrow 232.86.$$

$$\hat{y} = 1278 + 232.86x.$$

5) Filtering

In time series filtering is used to extract /isolate a particular component of the data set. The underlying signal while removing noise /unwanted components (trends /seasonal effects)

- It can help in
  1) Smoothing the data
  2) Removing trends / Detrending
  3) Isolating cyclical /seasonal components
  4) Enhancing signal to noise ratio.

Types of filters:

- There several types of filters used in T.S.

1) Moving Average filters

- In this filter each data point is replaced in the average of its neighbours. For a centered moving average with a window size $2k+1$, the M.A. filter is given by

$$\bar{y}_t = \frac{1}{(2k+1)} \sum_{i=k}^{k} y_{t+i} \ ---$$

smooth short term fluctuations and highlight long term trends

$\bar{y}_t$ = filtered value at time t

k = determines the window size

$2k+1$ = is the total window size.

- The aim of the MA filtering is to smooth short term flactuations and highlight long term trend.

Example.

Given the following data set

$y = (10, 12, 14, 16, 18, 20, 22)$

Apply a 3 point Moving Average Filter $k=1$

Solution

$t=3$

$\bar{y}_3 = \frac{1}{3} \sum_{i=-1}^{1} y_{3+i}$ Since $i = -1, 0, 1$

$\bar{y}_3 = \frac{y_{3-1} + y_3 + y_4}{3}$

$= \frac{y_2 + y_3 + y_4}{3}$

Checking from the data set

$y_2 = 12$ $y_3 = 14$ $y_4 = 16$.

$\frac{12+14+16}{3} = 14$

2) When $t=4$

$\bar{y}_4 = \frac{1}{3} \sum_{i=-1}^{1} y_{4+i}$

$\bar{y}_4 = \frac{1}{3}(y_{4-1} + y_4 + y_5)$

$= \frac{1}{3}(14+16+18)$

$= 16$

3) When $t=5$

$\bar{y}_5 = \frac{1}{3}(y_{5-1} + y_5 + y_6)$

$= \frac{1}{3}(16+18+20)$

$= \frac{1}{3}(54)$

$= 18$

4) $t=3$ $k=2$ $i=-2,-1,0,1,2$

$$\bar{y}_3 = \frac{1}{2(2)+1} \sum_{i=-2}^{2} \left(y_{3-2} + y_{3-1} + y_{3+0} + y_{3+1} + y_{3+2}\right)$$

$$= \frac{1}{5}\left(y_1 + y_2 + y_3 + y_4 + y_5\right)$$

$$= \frac{1}{5}\left(10+12+14+16+18+\cancel{20}+\cancel{22}\right) = \frac{1}{5}(70)$$

$$= 14.$$

5) $t=4$ $k=2.$

$$\bar{y}_4 = \frac{1}{2(2)+1} \sum_{i=-2}^{2} \left(y_{4-2} + y_{4-1} + y_{4+0} + y_{4+1} + y_{4+2}\right)$$

$$= \frac{1}{5}\left(y_2 + y_3 + y_4 + y_5 + y_6\right)$$

$$= \frac{1}{5}\left(12+14+16+18+20\right)$$

$$= \frac{1}{5}(80)$$

$$= 16.$$

6) $t=5$ $k=2.$

$$\bar{y}_5 = \frac{1}{2(2)+1} \sum_{i=-2}^{2} \left(y_{5-2} + y_{5-1} + y_{5+0} + y_{5+1} + y_{5+2}\right).$$

$$= \frac{1}{5}\left(y_3 + y_4 + y_5 + y_6 + y_7\right)$$

$$= \frac{1}{5}(90) = 18.$$

2) Exponential Smoothing

- In this method more ~~react~~ recent observations are given more weight. The simple exponential smoothing formular.

$$S_t = \alpha y_t + (1-\alpha)S_{t-1} \quad ---$$

where

Exponential smoothing? | smooth the T.S while maintaining responsives to changes

S_t is the smooth value of time t.
α ($0 \le \alpha \le 1$) is the smoothing parameter.
The aim is to smooth the T.S data while maintaining responsives to changes

Example.
Given the same data set in the last data set previous example and an initial smooth value $S_1 = 10$, Apply simple exponential smoothing with $\alpha = 0.5$. Calculate the smooth value at $t = 3$.

The reversive formula is

Calculate S_2

$$S_t = \alpha y_t + (1-\alpha) S_{t-1}$$

S_{2-1}

Calculate S_2.

$$\alpha y_2 + (1-\alpha) S_1$$

$$S_2 = 0.5(12) + (1-0.5)10$$
$$= 11$$

Calculate S_3.

$$S_3 = \alpha y_3 + (1-\alpha) S_2$$
$$= 0.5(14) + (1-0.5)11$$
$$= 12.5$$

Calculate S_5 ∴ Calculate for S_4

$$S_5 = \alpha y_4 + (1-\alpha) S_3$$
$$= 0.5(16) + (1-0.5)12.5$$
$$= 14.25$$

hence S_5.

$$\alpha y_5 + (1-\alpha) S_4$$
$$0.5(18) + (1-0.5)14.25$$
$$= 16.125$$

3) Differentiating Method

$S_5 = \alpha y_5 + (1-\alpha) S_4.$

$= 0.5(18) + (0.5)(14.5) = 16.25.$

iii> Differencing Method.

...ing is a transformation applied to a TS to remove ...and achieve stationarity.

...s ~~t_t~~ X_t the 1st order difference is defined as

...order $\Delta X_t = X_t - X_{t-1}$

... differencing the series its still exhibiting trends/season... order ~~man~~ differencing may be applied; and it is given as

$$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1}.$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

$(x-1)^3$ $(x-1)($

$(x-1)(x^2+1)$ $\boxed{1}$

$x^3 + x - x^2 - 1$

$x^3 - x^2 + x - 1$

3rd order $\Delta^3 X_t = \Delta X_t - \Delta^2 X_t$

$$= \Delta X_t - (\Delta X_t - \Delta X_{t-1})$$

$$= \Delta X_t - 3X_{t-3} + 3X_{t-2} + X_{t-1}$$

Example.

$= a + b\cancel{t} + \varepsilon_t$, $\varepsilon_i \sim iid(0, \delta^2)$.

the first order difference.

sol. $\Delta X_t = X_t - X_{t-1}$

$\Delta X_t = a + bt + \varepsilon_t - [a + b(t-1) + \varepsilon_{t-1}]$

$= a + bt + \varepsilon_t - a - b(t-1) - \varepsilon_{t-1}$

$= \cancel{a} + \cancel{bt} + \varepsilon_t - \cancel{a} - \cancel{bt} + b - \varepsilon_{t-1}$

$= b + \varepsilon_t - \varepsilon_{t-1}$

Example.

$= a + bt + ct^2 + \varepsilon_t$ $\varepsilon_i \sim iid(0, \delta^2)$. Make it stationary

sol.

1st difference, if we still have t we apply 2nd diff.

1st difference $\Delta X_t = X_t - X_{t-1}$

$\Delta X_t = a + bt + ct^2 + \varepsilon_t - [a + b(t-1) + c(t-1)^2 + \varepsilon_{t-1}$

$a + bt + ct^2 - \varepsilon_t - a - b(t-1) + c(t^2 - 2t + 1) + \varepsilon_{t-1}$

$\cancel{a} + \cancel{bt} + \cancel{ct^2} - \varepsilon_t - \cancel{a} - \cancel{bt} + b + \cancel{ct^2} + 2ct - c + \varepsilon_{t-1}$

$\varepsilon_t + 2ct - c + \varepsilon_{t-1} + b.$

$b - c + 2ct + \varepsilon_t + \varepsilon_{t-1}$

2nd difference $\Delta^2 X_t = \Delta X_{t-1}$ $\quad 2c + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}$

$= X_t - 2X_{t-1} + X_{t-2}.$

$b - c + 2ct + \varepsilon_t + \varepsilon_{t-1} - 2(b - c + 2ct + \varepsilon_t + \varepsilon_{t-1}) + b + c + 2ct_{-2} + \varepsilon_{t-2} + \varepsilon_{t-3}.$

$\cancel{b} - \cancel{c} + 2ct + \varepsilon_t + \varepsilon_{t-1} - \cancel{2b} - \cancel{2c} - 4ct - 2\varepsilon_t - 2\varepsilon_{t-1}\} + \cancel{b} + \cancel{c} + 2ct_{-2} + \varepsilon_{t-2} + \varepsilon_{t-3}.$

Mathematical Models for Time Series:

There are 2 models mostly used for the decomposition of a TS. These are additive model and multiplicative model.

i> Additive Model

It assumes that the observed value is the sum of the 4 components of the TS: This is given as

$$X = T + S + C + I$$

where T = Trend ~~Value~~ Value

S = Seasonal component

C = Cyclic component

I = Irregular component

X = Observed component

ii> Multiplicative

It assumes that the observed value is obtained by multiplying the values of the 4 components. It's given as

$$X = T \times S \times C \times I.$$

The definitions are still the same.