module · 1

diff · between regression and LSA.

PAGE No.
DATE / /

* What is time series ?

→ A set of numeric values of some variable obtained at regular period over time.

eg. daywise, monthwise, year wise, etc.

* Difference between regression and time series.

→

Regression	Time series
i. It is to use some quantity to predict some other quantity within a given range, i.e. relationship between two variables.	i. Time series is a relationship between lagged version of the data and the predicting series.
ii. Data points are independent.	ii. Depends on the value given but also on the sequence in which the values are collected.

* Difference between interpolation and extrapolation.

→

Interpolation	Extrapolation
i. Predicting within the given range. For example: If temperature is given and we need to find the current temperature.	i. Extrapolation is predicting outside the given range. For example: If the temperature of past 2 days has been provided and the model should predict next couple of days.

* Difference between time series analysis and time series forecasting.

→ Time series Analysis	Time series Forecasting
i. TSA is understanding the dataset with respect to components such as trend, cycle, seasonality or irregularity.	i. It is applying an appropriate model to predict the future values based on the patterns obtained while analysis.

* Types of time series.

→ i. Classification - Identifies and assigns categories to dataset.

ii. Curve fitting - Plots the data along a curve to study the relationship of variables within the data.

iii. Descriptive analysis - Identifies patterns in the time series data based on the components.

iv. Explanative analysis - Attempts to understand the data, the relationship, the causes as well as the effect on the data.

v. Exploratory analysis - Highlights the main characteristics of the time series data in a visual format.

vi. Forecasting - Predicts future data based on historical (past) trends.

Q. which graph is irregular and which is noise?

avg. is ≈ 0 always.

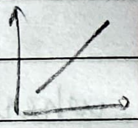
PAGE No.

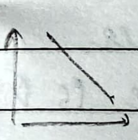
DATE / /

vii. Intervention analysis - study of how an event can change the data, i.e. comparing before intervention and after intervention.

* Types of time series patterns / components:

→ i. Secular trend / Trend (T) - A steady tendency of either an upward / downward movement in the average value of the forecast variable over time.



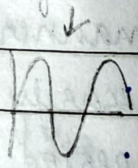


Short term period: Studying the effect of human pulse rate.

Long term period: National income, agricultural production, etc.

ii. Cyclic variations / Cycle (C) - An upward or downward movement in the variable value over the trend line with four phases -

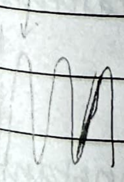
longer



- From peak (prosperity)
- Contradiction (recession)
- Depression (tough market situation)
- Expansion (growth / recovery)

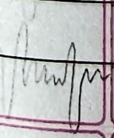
iii. Seasonal variations / Seasonal (S) - A special case of cyclic component where fluctuations are repeated usually within a time frame.

shorter



Examples are: sales of stationary items increases at the start of an academic year, sales of ice-cream increases in summers.

iv. Irregular / Residual variations (I/R) - Also known as random pattern i.e. characterized by unpredictable variations that do not follow a



Q. If difference is asked, last point should be graphs.

PAGE No.
DATE / /

consistent or repetitive structure over time. This pattern might still exhibit some level of correlation or dependency across different time points.

* White Noise - • It is a specific type of time series that consists of random, un-correlated values with a constant mean and variance.
• It represents a complete random process where each observation is independent of each other.
• The average value of white noise over time is 0.

* Difference between seasonal and oscillatory pattern.

Seasonal Pattern	Oscillatory Pattern
i. Regular and predictable fluctuations in a time series that occur at specific fixed periods. Eg. daily, monthly, weekly	i. Refers to fluctuations in a time series that move in a wave like manner but are not necessarily tied to a fixed period or calender events.
ii. Predictable pattern with known frequency.	ii. Amplitude and frequency of oscillation can vary making it less predictable.

* Models of time series analysis :
1. Additive model (used when seasonal variations are constant)
$Y(t) = T(t) + S(t) + C(t) + I(t)$

2. Multiplicative model (seasonal variations are increasing)
$Y(t) = T(t) \times S(t) \times C(t) \times I(t)$

- moving average (window size: 3)
- least squares will come in TT or end sem compulsary (5-8 mks)
- always label (- - - ← actual values, — ← trend line)

PAGE No. | DATE / /

* Measurement of trends -

only check the trend.
1. Freehand or graphic method (in ppt)
2. Moving Average method → trend line acc. to avg.
3. Semi- Average method

forecast the trend.
4. Least squares.

- Method of least squares - The line of best fit is a line which the sum of deirivation is 0 of various points.
It is the best method to obtain the trend values.
The sum of deirivation of actual values (Y) and predicted / forecast value ($\hat{Y}$) is 0.

Ex. 1 (odd years) For the following time series fit a linear trend by the method of least squares. Also find the trend values and estimate the sales for the year 2017.

Year	Sales(Y) (1000)	X	XY	X^2	Trend value (1000)
2010	125	-3	-375	9	125.679
2011	128	-2	-256	4	128.786
2012	133	-1	-133	1	131.893
origin → 2013	135	0	0	0	0
2014	140	1	140	1	138.107
2015	141	2	282	4	141.214
2016	143	3	429	9	144.321
	945	0	87	28	

PAGE No.
DATE / /

only for odd ↓

→ Step 1: $X = \frac{\text{Years} - \text{Origin}}{\text{interval}}$

★ In exam, solve properly for 2 values, then directly write the ans.

time frame → $X_{2010} = \frac{2010 - 2013}{1} = -3$

$X_{2011} = \frac{2011 - 2013}{1} = -2$

$X_{2012} = \frac{2012 - 2013}{1} = -1$

$X_{2013} = \frac{2013 - 2013}{1} = 0$

$X_{2014} = \frac{2014 - 2013}{1} = 1$

$X_{2015} = \frac{2015 - 2013}{1} = 2$

$X_{2016} = \frac{2016 - 2013}{1} = 3$

Step 2: calculate: ΣY, ΣX, ΣXY, ΣX^2

$\Sigma Y = 945$

$\Sigma X = 0$

$\Sigma XY = 87$

$\Sigma X^2 = 28$

$Y = a + bx$ → equation of straight line trend

similarly, $\hat{Y} = \underline{a + bx}$ → (constant)

Step 3:

simultaneous eq → $\begin{cases} \Sigma Y = Na + b\Sigma x \\ \Sigma XY = a\Sigma x + b\Sigma x^2 \end{cases}$ → where N = no. of years

i. $945 = 7a + b(0) \Rightarrow a = 135$

compulsary 3 decimal places ⇒ ii. $87 = 135(0) + b(28) \Rightarrow b = 3.107$

FOR 6-7 mks solve like this].

FOR 10 mks, show calculation for each step.

PAGE No.
DATE / /

Step 4: Calculate the trend values:

$\hat{y}_{2010} = a + bX = 135 + 3.107(-3) = 125.679$

$\hat{y}_{2011} = a + bX = 135 + 3.107(-2) = 128.786$

$\hat{y}_{2012} = 131.893$

$\hat{y}_{2013} = 0$

$\hat{y}_{2014} = 138.107$

$\hat{y}_{2015} = 141.214$

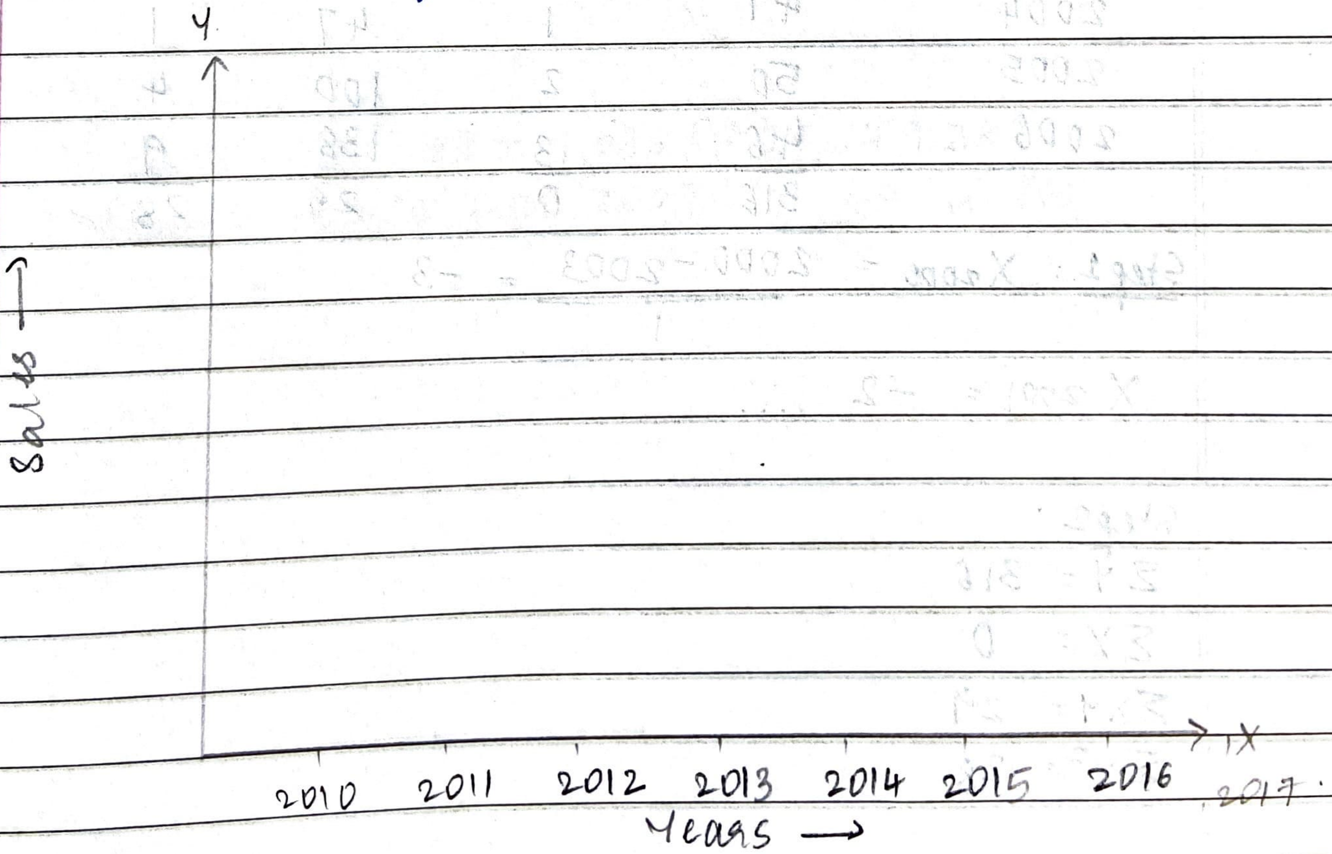
$\hat{y}_{2016} = 144.321$

Step 5: calculate for 2017

$X = \frac{2017 - 2013}{1} = 4$

$\therefore \hat{y}_{2017} = 135 + 3.107(4) = \underline{147.428}$

Estimated sales for 2017 is ₹147428.



PAGE No.

DATE / /

Ex 2 Given below are the data relating to production of sugarcane in a district. Fit a straight line trend by the method of least squares. Tabulate the trend values and estimate for the year 2007.

Year	2000	2001	2002	(2003)	2004	5	6
Prod. of sugarcane	40	45	46	42	47	50	46

→

Year	Prod. (Y)	X	XY	X^2	Trend val.
2000	40	-3	-120	9	42.035
2001	45	-2	-90	4	43.071
2002	46	-1	-46	1	44.107
(2003)	42	0	0	0	45.143
2004	47	1	47	1	46.179
2005	50	2	100	4	47.215
2006	46	3	138	9	48.251
	316	0	29	28	

Step 1: $X_{2000} = \frac{2000-2003}{1} = -3$

$X_{2001} = -2$ ....

Step 2:

$\Sigma Y = 316$

$\Sigma X = 0$

$\Sigma XY = 29$

$\Sigma X^2 = 28$

$y = a\cancel{x} + bx$

similarly $\hat{y} = a + bx$

Step3 :

$\Sigma Y = Na + b\Sigma x$

$\Sigma XY = a\Sigma x + b\Sigma x^2$

$\therefore$ $316 = 7a + b(0) \Rightarrow a = 45.143$

$\therefore$ $29 = 45.143(0) + b(28) \Rightarrow b = 1.036$

Step4 :

$\hat{Y}_{2000} = a + bX = 45.143 + 1.036(-3) = 42.035$

$\hat{Y}_{2001} = 43.071$

$\hat{Y}_{2002} = 44.107$

$\hat{Y}_{2003} = 45.143$

$\hat{Y}_{2004} = 46.179$

$\hat{Y}_{2005} = 47.215$

$\hat{Y}_{2006} = 48.251$

Step5:

$\hat{Y}_{2007} = 45.143 + 1.036(4) = 49.287$

Estimated sales for 2007 is 49 units.

$\left\{ \frac{\text{Year} - \text{Arithmetic mean of 2 middle years}}{\text{Half of interval}} \right\}$

PAGE No.

DATE / /

EX3 For the following time series fit a linear trend by the method of least squares. Find the trend values and estimate the sales for the year 2017.

even no. of years

Year	Value(Y)	X	XY	X^2	Trend Values
2009	80	-7	-560	49	
2010	90	-5	-450	25	
2011	92	-3	-276	9	
2012	83	-1	-83	1	
2013	94	1	194	1	
2014	99	3	297	9	
2015	92	5	460	25	
2016	104	7	728	49	
	734	0	210	168	

→ Step 1 : For even no. of years, $X = \frac{\text{Year} - \text{AM of 2 mid yrs}}{\text{Half of interval}}$

AM of 2 middle years (origin) $= \frac{2012 + 2013}{2} = 2012.5$

$X_{2009} = \frac{2009 - 2012.5}{0.5(1)} = -7$

$X_{2010} = \frac{2010 - 2012.5}{0.5(1)} = -5$

$X_{2011} = -3 \ldots$

$X_{2012} = = -1$

$X_{2013} = 1$

$X_{2014} = 3$

$X_{2015} = 5$

$X_{2016} = 7$

PAGE No.
DATE / /

Step 2:

$\Sigma Y = 734$

$\Sigma X = 0$

$\Sigma XY = 210$

$\Sigma X^2 = 168$

$\hat{y} = a + bx$

Step 3:

$\Sigma Y = Na + b\Sigma x \rightarrow (1)$

$\Sigma XY = a\Sigma x + b\Sigma x^2 \rightarrow (2)$

$\therefore 734 = 8a + b(0) \Rightarrow a = 91.75$

$\therefore 210 = 0 + b(168) \Rightarrow b = 1.25$

Step 4:

$\hat{y} = a + bx$

$\therefore \hat{y}_{2009} = 83$

$\hat{y}_{2010} = 85.5$

$\hat{y}_{2011} = 88$

$\hat{y}_{2012} = 90.5$

$\hat{y}_{2013} = 93$

$\hat{y}_{2014} = 95.5$

$\hat{y}_{2015} = 98$

$\hat{y}_{2016} = 100.5$

PAGE No.
DATE / /

EX.4 Given below is the data relating to the sales of a product in a district. Fit a straight line trend by the method of least squares, tabulate the trend values and estimate for the year 2003.

Year	Sales(Y)	X	XY	X^2	Trend val
1995	6.7	-7	-46.9	49	5.618
1996	5.3	-5	-26.5	25	5.72
1997	4.3	-3	-12.9	9	5.822
1998	6.1	-1	-6.1	1	5.924
1999	5.6	1	5.6	1	6.026
2000	7.9	3	23.7	9	6.128
2001	5.8	5	29	25	6.23
2002	6.1	7	42.7	49	6.332
	47.8	0	8.6	168	

→ Step 1 : $AM = \frac{1998 + 1999}{2} = 1998.5$

$X_{1995} = \frac{1995 - 1998.5}{0.5} = -7$

$X_{1996} = -5$ . . . .

Step 2 : $\Sigma Y = 47.8$, $\Sigma X = 0$, $\Sigma XY = 8.6$, $\Sigma X^2 = 168$

Step 3 : $\Sigma Y = Na + b\Sigma x$

$\Sigma XY = a\Sigma x + b\Sigma x^2$

$\therefore 47.8 = 8a + 0 \Rightarrow a = 5.975$

$\therefore 8.6 = a.0 + b(168) \Rightarrow b = 0.051$

Step 4 .

$\hat{y} = a + bx$

$\hat{Y}_{1995} = 5.618$ . . . .

Step 5:

For 2003:

$X = 9$

$\hat{Y}_{2003} = 6.434$

$\therefore$ Estimated sales for the year 2003 is ₹ 6.434.

PAGE No.
DATE / /

* Seasonality :
Seasonality in time series refers to a pattern or fluctuation that repeats at regular intervals within a fixed time frame.
Types of seasonality :
1. Additive seasonality : $\cdot Y = T(t) + C(t) + S(t) + I(t)$.
· values are independent of each other.
2. multiplicative seasonality: $\cdot Y = T(t) \times C(t) \times S(t) \times I(t)$
· values are dependent of each other.

Detect and analyse seasonality in time series :
1. Visual inspection - Examining the pattern for any regular and repetitive fluctuations (plotting the variable to be analyzed on a graph).

2. Autocorrelation function (ACF) and partial autocorrelation function (PACF) - Statistical tools can be used to identify the presence of seasonality by examining the correlation between the time series observation and the lagged values.

3. Decomposition - Time series decomposition technique such as additive or multiplicative decomposition can separate a time series into various time series components to enable the presence of seasonality.

PAGE No.
DATE / /

* Seasonal Indices - 8.

- Seasonal variation is measured in terms of an index called as seasonal indices.
- It is an average that can be used to compare an actual observation relative to what it would be if there were no seasonal variation.

NOTE: All seasonal index average should be 1.
- Monthly : add to 12
- Quarterly : add to 4

EX.1

Month	Seasonal Index	Sales (1000)
Jan	1.2	72
Feb	1.2	72
~~March~~	1.1	66
Apr	1	60
May	0.95	57
Jun	0.95	57
Jul	0.9	54
Aug	0.85	51
Sept	0.85	51
Oct	0.9	?
Nov	1	60
Dec	?	61

1. Find December's seasonal index.

→ Add all seasonal indices: $10.9 + x$

$\therefore \frac{10.9+x}{12} = 1 \Rightarrow x = 1.1$

$\therefore$ December's seasonal index is 1.1.

The meaning of seasonal index in the above equation

PAGE No.
DATE / /

is 1.1 which is 10% above average.

* Deseasonalizing data:

Deseasonalizing figure = $\frac{\text{Actual figure}}{\text{seasonal index}}$

Example 1: The actual sales of a company for the month of August was \$51000 and the seasonal index was 0.85. Find the deseasonalized sales.

→ deseasonalized figure = $\frac{51000}{0.85}$ = \$60000

* Reseasonalizing ~~fac~~ data:

Actual figure = Deseasonalized figure × SI

EX.1 The deseasonalized sale for October was \$60000 with an SI of 0.85. What were the actual sales?

→ actual figure = 60000 × 0.85 = \$51000

* Deseasonalized data in percentage.

EX. By what percentage do we reduce January's sales of \$72000 to deseasonalize? SI = 1.2

→ deseasonalize = Actual × $\frac{1}{SI}$

= ~~72000~~ 72000 × $\frac{1}{1.2}$ = \$60000

~~By~~ To what % it got reduced = ~~\$~~ 83.33%

∴ Reduced by: 100 − 83
= 17%

PAGE No.
DATE / /

Q. compute the seasonal index of the following data:

Quarter	Summer	Autumn	Winter	Spring
Sales	920	1085	1241	446

i. Find the seasonal average of the data given.
ii. Find the seasonal index for each quarter.

→ i. Seasonal average = $\frac{920 + 1085 + 1241 + 446}{4} = 923$

ii ~~For summer:~~
Seasonal index = $\frac{\text{Value of individual season}}{\text{seasonal average}}$

- For summer, $SI = \frac{920}{923} = 0.9976$ don't round up/down the last digit
- For Autumn, $SI = \frac{1085}{923} = 1.1765$
- For winter, $SI = \frac{1241}{923} = 1.3454$
- For spring, $SI = \frac{446}{923} = 0.483$

Q. Compute the seasonal indices of the foll. data:
i. Find seasonal average of each year.
ii. Find seasonal indices of individual season of each year in respective quarter.
iii. Calculate the overall seasonal indices of each quarter.
iv. Deseasonalize the data of each year in respective quarter.

PAGE No.

DATE / /

Year	Summer	Autumn	Winter	Spring
1	920	1085	1241	446
2	1035	1180	1356	541
3	1299	1324	1450	659

→ i. avg. $SI_1 = \frac{3692}{4} = 923$

avg. $SI_2 = \frac{4112}{4} = 1028$

avg. $SI_3 = \frac{4732}{4} = 1183$

⇒

Year	Seasonal Avg.
1	923
2	1028
3	1183

ii.

~~SI~~ Year	SI_{summer}	$SI_{Aut.}$	$SI_{wint.}$	$SI_{spr.}$
1	$\frac{920}{923} = 0.996$	$\frac{1085}{923} = 1.175$	$\frac{1241}{923} = 1.344$	$\frac{446}{923} = 0.483$
2	$\frac{1035}{1028} = 1.006$	$\frac{1180}{1028} = 1.147$	$\frac{1356}{1028} = 1.319$	$\frac{541}{1028} = 0.526$
3	$\frac{1299}{1183} = 1.\cancel{263}$ 098	$\frac{1324}{1183} = 1.119$	$\frac{1450}{1183} = 1.225$	$\frac{659}{1183} = 0.55$

iii. seasonal avg (summer) $= \frac{920 + 1035 + 1299}{3} = 1084.666$

seasonal avg (aut.) $= \frac{1085 + 1180 + 1324}{3} = 1196.333$

Not required

seasonal avg (wint.) $= \frac{1241 + 1356 + 1450}{\cancel{4}\ 3} = 1349$

seasonal avg. (spring) $= \frac{446 + 541 + 659}{3} = 548.666$

Year	SI sum.	SI aut.	SI wint.	SI spr.
1	0.996	1.175	1.344	0.483
2	1.006	1.147	1.319	0.526
3	098 1.263	1.119	1.225	0.557
overall =	$\frac{3.1[illegible]}{3} = 1.033$	$\frac{3.441}{3} = 1.147$	$\frac{3.888}{3} = 1.296$	$\frac{1.566}{3} = 0.5$

previous values.

iv.

Q. Calculate the seasonal index for the quarterly production of a product using the method of simple averages.

Year	I^{st} Quarter	II^{nd} Quarter	III^{rd} Quart.	IV^{th} Quar.
2005	255	351	425	400
2006	269	310	396	410
2007	291	332	358	395
2008	198	289	310	357
2009	200	290	331	359
2010	250	300	350	400
Quarterly Total	1463	1872	2170	2321
Quarterly Average	243.833	312	361.666	386.83

$$\text{Grand average} = \frac{\text{Total of quarterly averages}}{n} = 326.083$$

$$\text{SI Quarter I} = \frac{243.833}{326.083} = 0.747$$

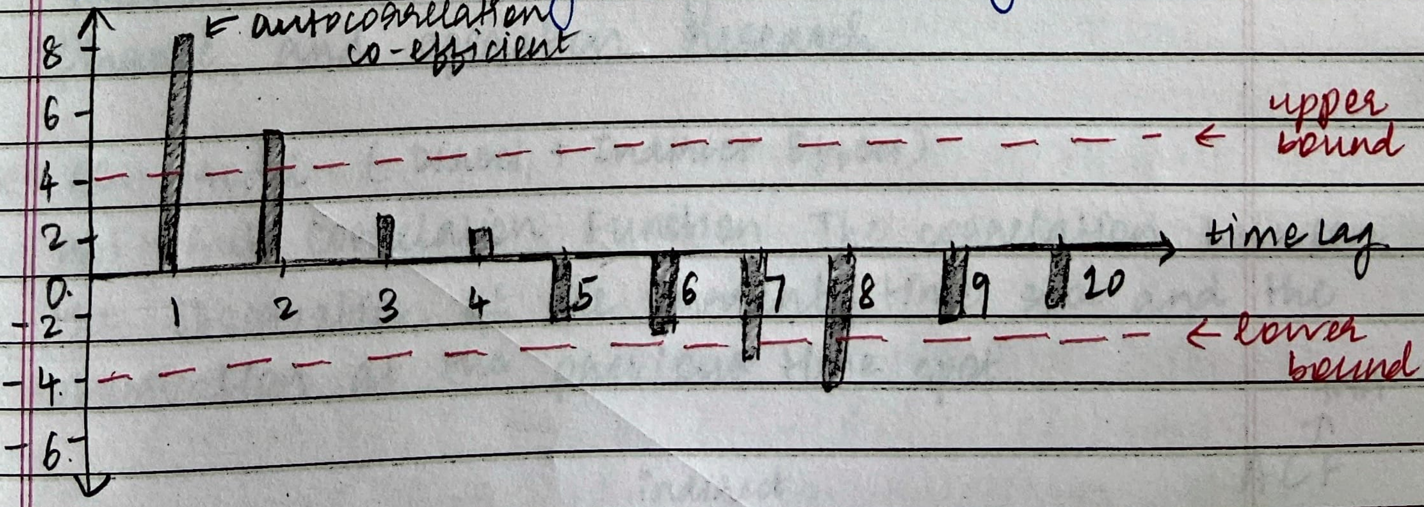
$$\text{SI Quarter II} = \frac{312}{326.083} = 0.956$$

$$\text{SI Quarter III} = \frac{361.666}{326.083} = 1.109$$

$$\text{SI Quarter IV} = \frac{386.833}{326.083} = 1.184$$

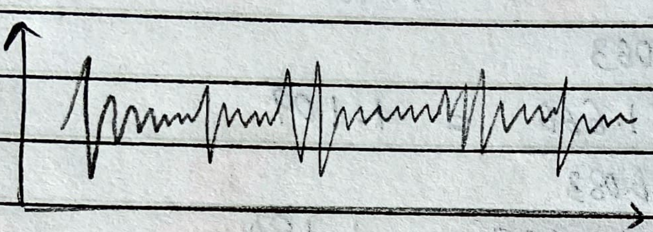
* Stochastic processes -
It is a collection of random variables indexed by time, where the values of the variables at each time point are determined by a probability distribution. In simpler terms, it is a sequence of random events or observations that unfold over time.

* Correlogram - The plot of auto-correlation function (ACF) versus time lag is called correlogram.

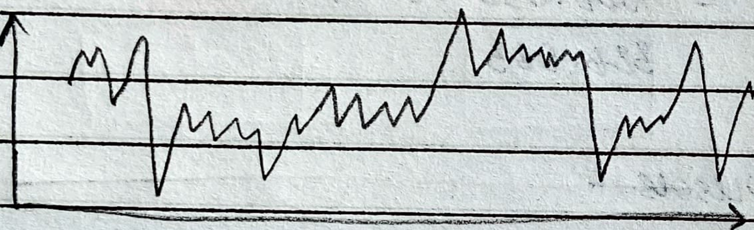


* Stationary Time Series -

Stationary means that the statistical properties of a process generating a time series do not change over time.



Stationary



Non-stationary