TIME SERIES DATA ANALYSIS

Kumar Sundram

TIME SERIES ANALYSIS

- A time series is a sequence of numerical data points in successive order.
- E.g. In investing, a time series tracks the movement of the chosen data points, such as a security's price, over a specified period of time with data points recorded at regular intervals.
- There is no minimum or maximum amount of time that must be included, allowing the data to be gathered in a way that provides the information being sought by the investor or analyst examining the activity.
- According to classical time-series analysis an observed time series is the combination of some pattern and random variations.
- The aim is to separate them from each other in order to
    - describe to historical pattern in the data,
    - prepare forecasts by projecting the revealed historical pattern future.

EXAMPLES

1. Real gross national product in the United States, 1872-1985 (annual).
2. Rate of growth of real gross national product in the United States, 1873-1985 (annual).
3. Unemployment rate in the United States, 1873-1985 (annual).
4. Price level in the United States, 1870-1985 (annual).
5. Inflation rate in the United States, 1870-1985 (annual).
6. Logarithm of retail sales of men's and boys' clothing in the United States, 1967-1979 (quarterly).

TYPES OF TIME SERIES ANALYSIS

Time series analysis can be used to accomplish different goals:

- Descriptive analysis determines what trends and patterns a time series has by plotting or using more complex techniques.

- The most basic approach is to graph the time series and look at:

  - Overall trends (increase, decrease, etc.)
  - Cyclic patterns (seasonal effects, etc.)
  - Outliers – points of data that may be erroneous
  - Turning points – different trends within a data series

TYPES OF TIME SERIES ANALYSIS

Spectral analysis is carried out to describe how variation in a time series may be accounted for by cyclic components.

This may also be referred to as "Frequency Domain".

With this an estimate of the spectrum over a range of frequencies can be obtained and periodic components in a noisy environment can be separated out.

Example: What is seen in the ocean as random waves may actually be a number of different frequencies and amplitudes that are quite stable and predictable. Spectral analysis is used on the wave height vs. time to determine which frequencies are most responsible for the patterns that are there, but can't be readily seen without analysis.

TYPES OF TIME SERIES ANALYSIS

Spectral analysis is carried out to describe how variation in a time series may be accounted for by cyclic components.

This may also be referred to as "Frequency Domain".

With this an estimate of the spectrum over a range of frequencies can be obtained and periodic components in a noisy environment can be separated out.

Example: What is seen in the ocean as random waves may actually be a number of different frequencies and amplitudes that are quite stable and predictable. Spectral analysis is used on the wave height vs. time to determine which frequencies are most responsible for the patterns that are there, but can't be readily seen without analysis.

UNIVARIATE TIME SERIES

A univariate time series is a sequence of measurements of the same variable collected over time. Most often, the measurements are made at regular time intervals.

difference from standard linear regression

- the data are not necessarily independent and not necessarily identically distributed.
- the ordering matters.

CONCEPTUAL TOPICS

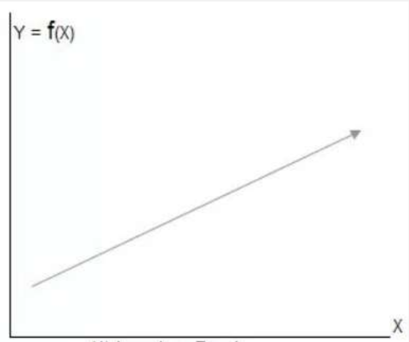
Topics	Description
Stationarity	A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time.
Trends	shows the general tendency of the data to increase or decrease during a long period of time.
Seasonality	seasonality is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly, or quarterly.
Cyclical movements	A cyclic pattern exists when data exhibit rises and falls that are not of fixed period.
Unexpected variations	These fluctuations are unforeseen, uncontrollable, unpredictable, and are erratic. These forces are earthquakes, wars, flood, famines, and any other disasters.

Traditional methods of time series analysis are concerned with decomposing of a series into a trend, a seasonal variation and other irregular fluctuations.

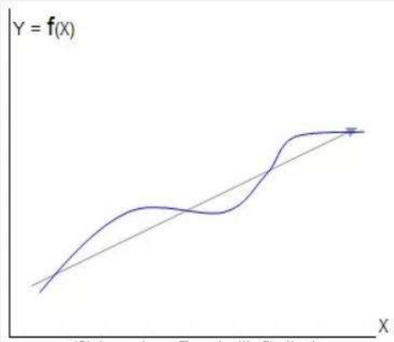
TREND

- a general systematic linear or (most often) nonlinear component that changes over time and does not repeat

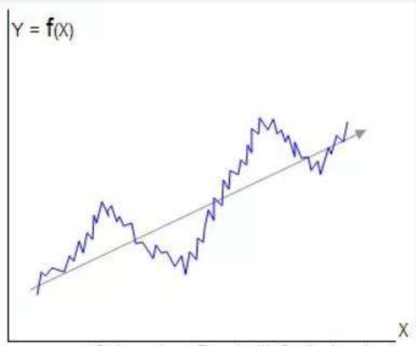
- Secular Trend or Long Term Variation (can be linear, non-linear, i.e. exponential, quadratic



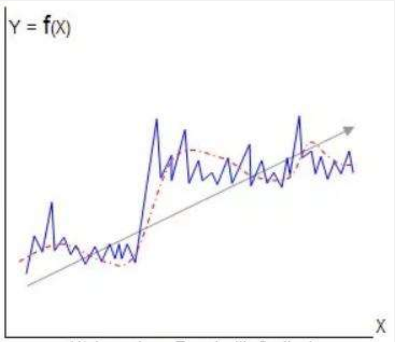
Long term trend



Long term trend with cyclical variations



Long term trend with cyclical and seasonal variations



Long term trend with cyclical, seasonal and random variations

TRENDS - TYPES

2 general classes of trends:

A time series with a trend is called **non-stationary.**

- Deterministic Trends: These are trends that consistently increase or decrease.
- Stochastic Trends: These are trends that increase and decrease inconsistently.

In general, deterministic trends are easier to identify and remove

Another way to look at trends:-

- Global Trends: These are trends that apply to the whole time series.
- Local Trends: These are trends that apply to parts or subsequences of a time series.

Generally, global trends are easier to identify and address.

SEASONALITY

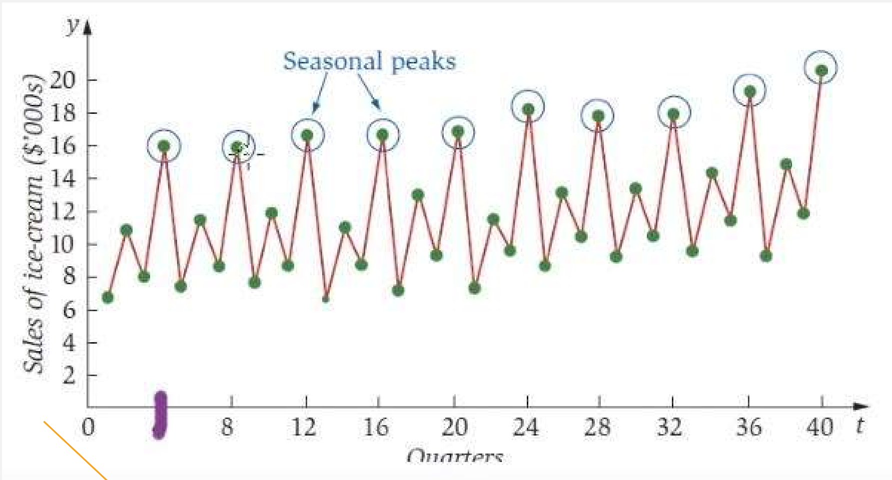
- seasonality is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly, or quarterly. (no longer than a year)

- may be caused by various factors, such as weather, vacation, and holidays and consists of periodic, repetitive and generally regular and predictable patterns (peaks and troughs)

- periodic time series - Seasonality is always of a fixed and known period.

- Repetitive

- Does not have to tie to seasons!



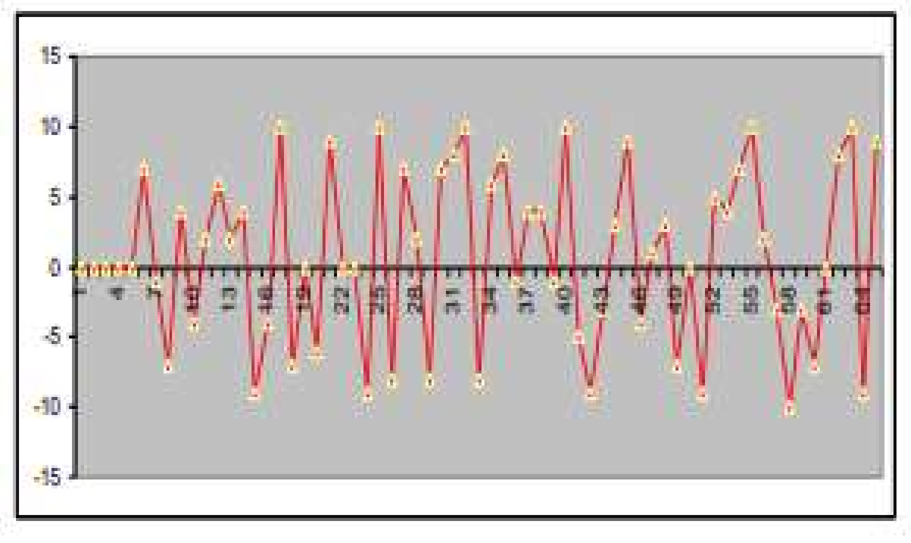
1. Seasonal
2. With positive secular trend

CYCLICAL MOVEMENTS

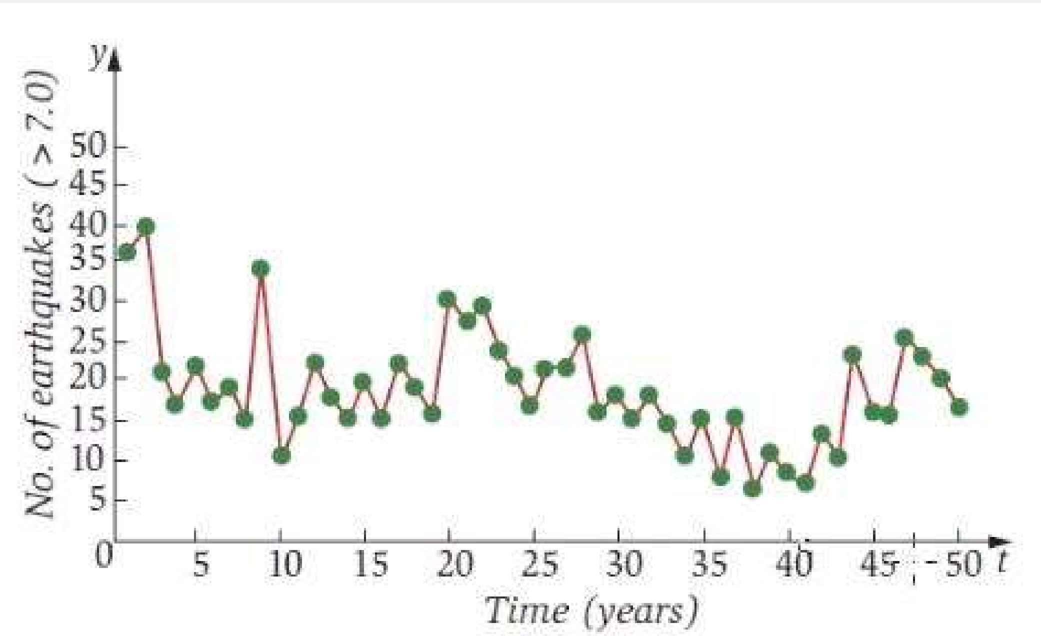
- A cyclic pattern exists when data exhibit rises and falls that are not of fixed period.
- unpredictable
- The duration of these fluctuations is usually of at least 2 years.
- Think of business cycles which usually last several years, but where the length of the current cycle is unknown beforehand.
- no set amount of time between those fluctuations.

UNEXPECTED VARIATIONS

- The random variations of the data comprise the deviations of the observed time series from the underlying pattern.

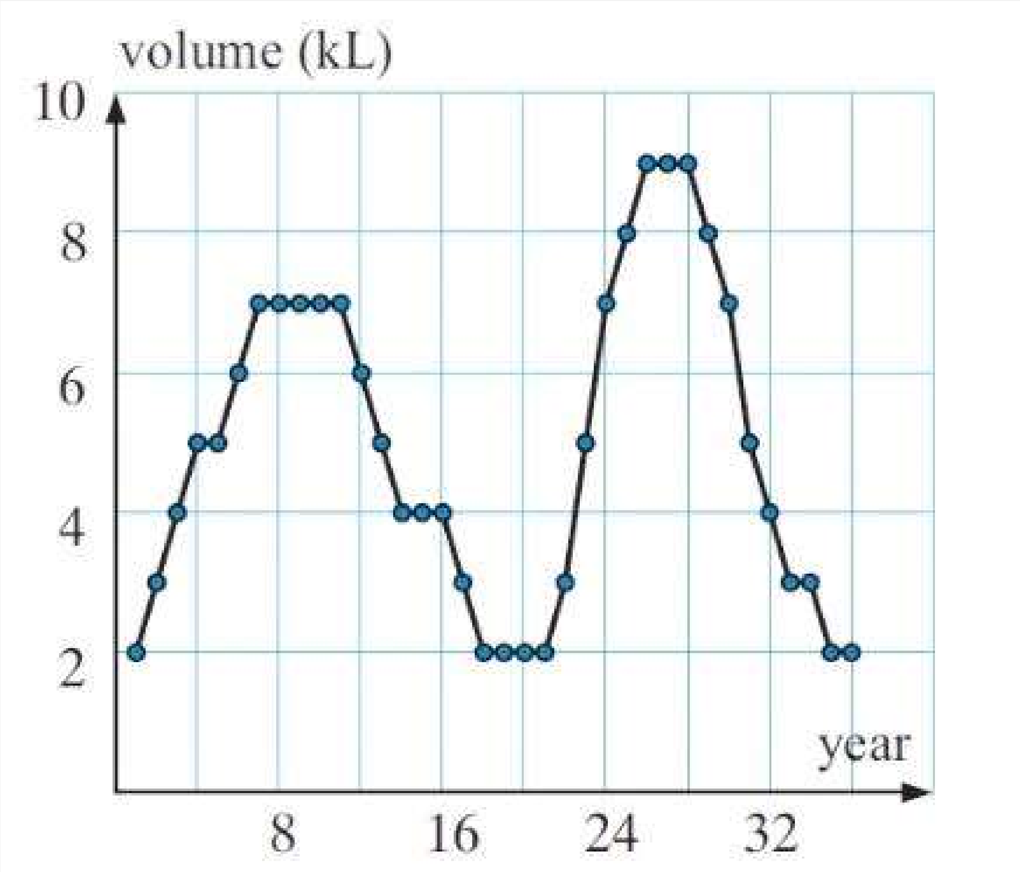


EXAMPLE



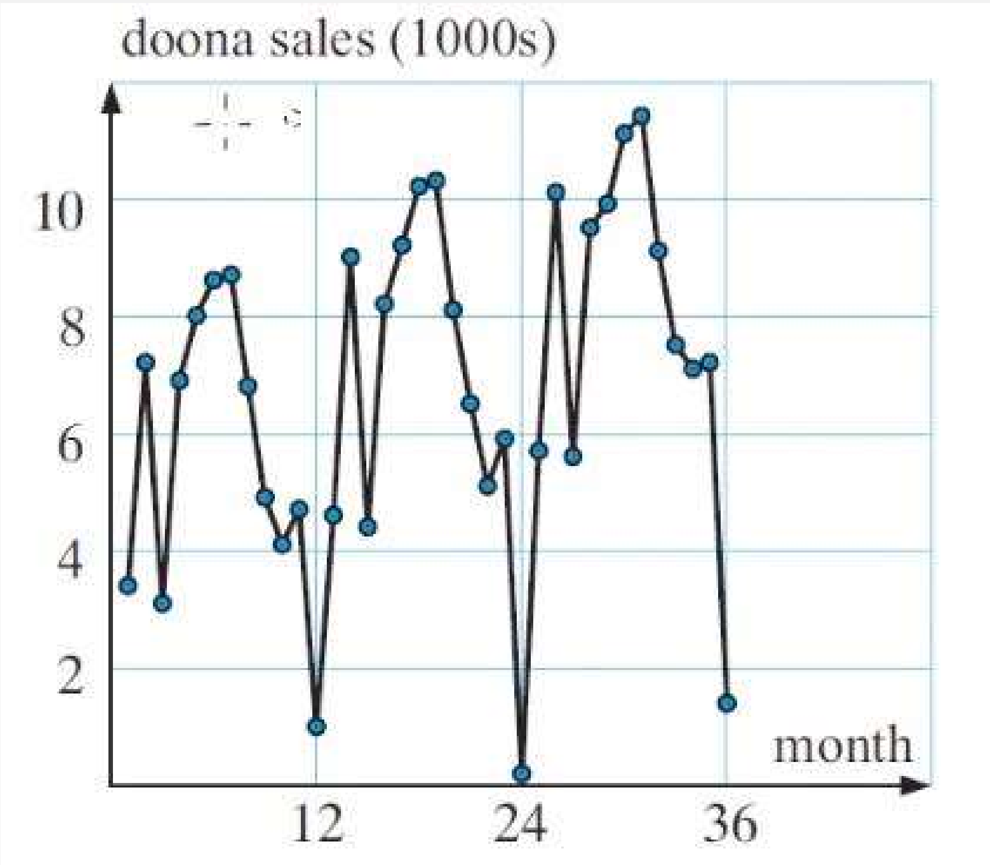
Aspect	comments
Seasonality	Peaks **not** regular periodicity
Trend	No apparent trend
Cyclic	Yes, as the peaks and troughs are irregular
Random	

EXAMPLE



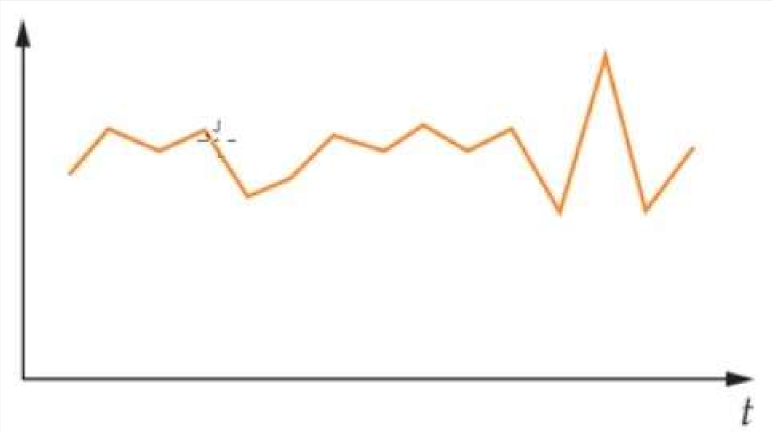
Aspect	comments
Seasonality	Peaks **not** regular periodicity
Trend	Slight positive secular trend
Cyclic	Yes, as the peaks and troughs are irregular
Random	

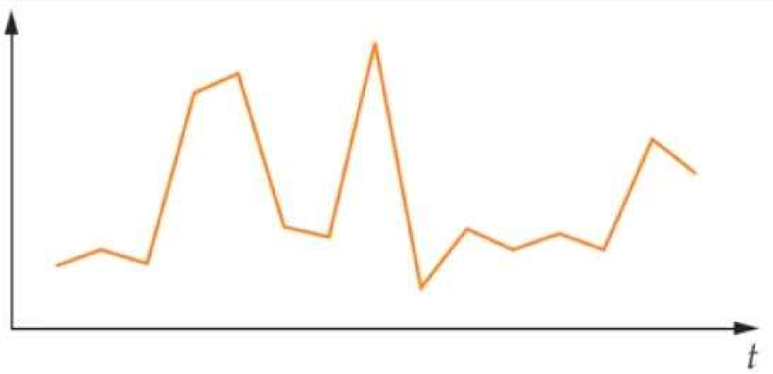
EXAMPLE



Aspect	comments
Seasonality	Yes, Peaks and troughs occur at regular periodicity
Trend	Slight positive/upward secular trend
Cyclic	No
Random	No

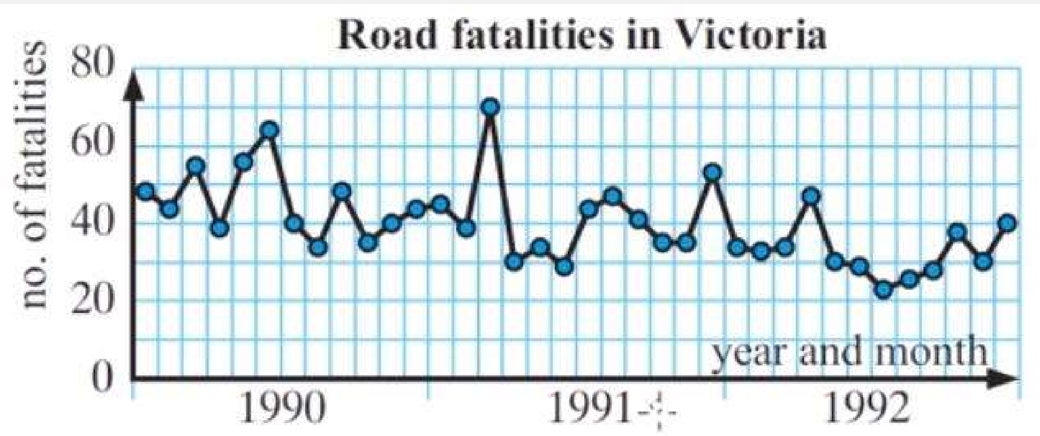
EXAMPLE –





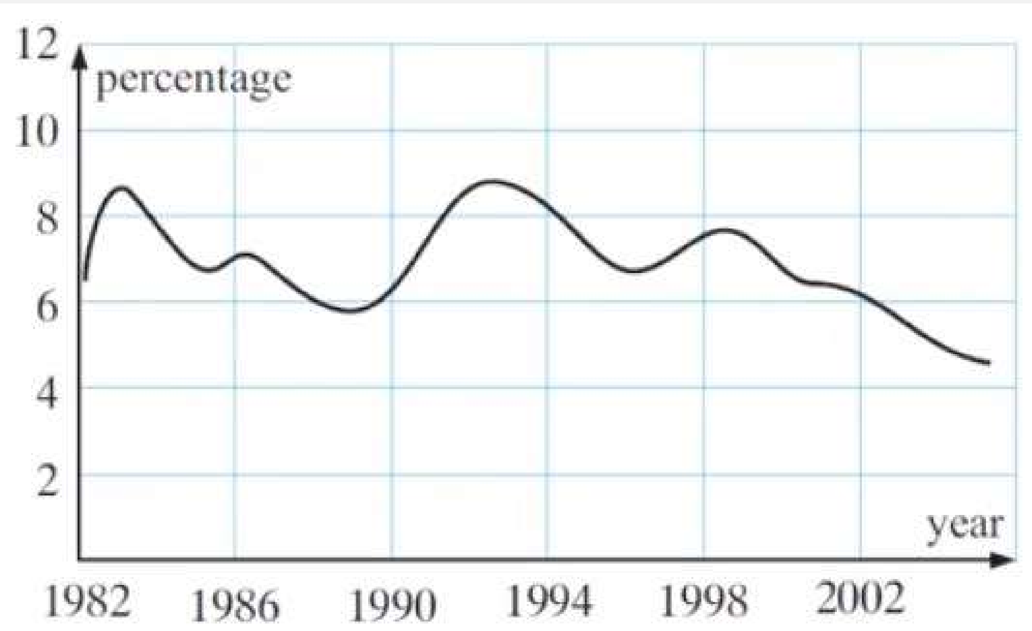
Aspect	comments
Seasonality	no
Trend	possibly
Cyclic	No
Random	Yes

EXAMPLE – RANDOM TIME SERIES



Aspect	comments
Seasonality	No
Trend	Downward negative trend
Cyclic	No
Random	Yes

EXAMPLE

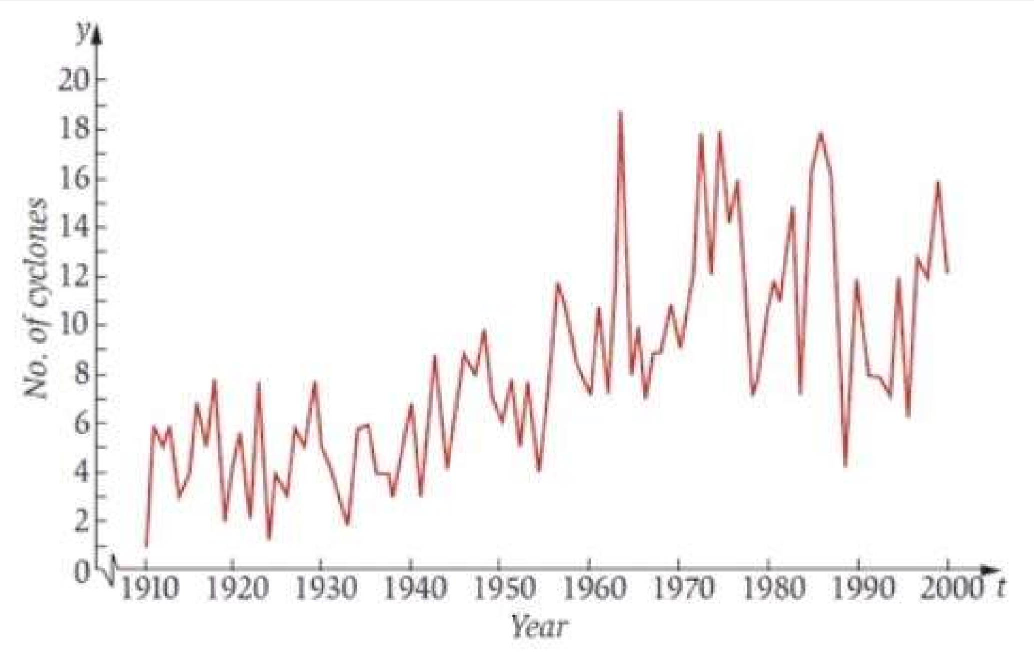


Aspect	comments
Seasonality	No
Trend	Slight Downward negative trend
Cyclic	Yes (peaks at irregular points)
Random	No

EXAMPLE

Aspect	comments
Seasonality	Yes
Trend	No trend
Cyclic	No
Random	No

EXAMPLE



Aspect	comments
Seasonality	No
Trend	Upward secular trend
Cyclic	Yes (peaks at irregular points)
Random	No

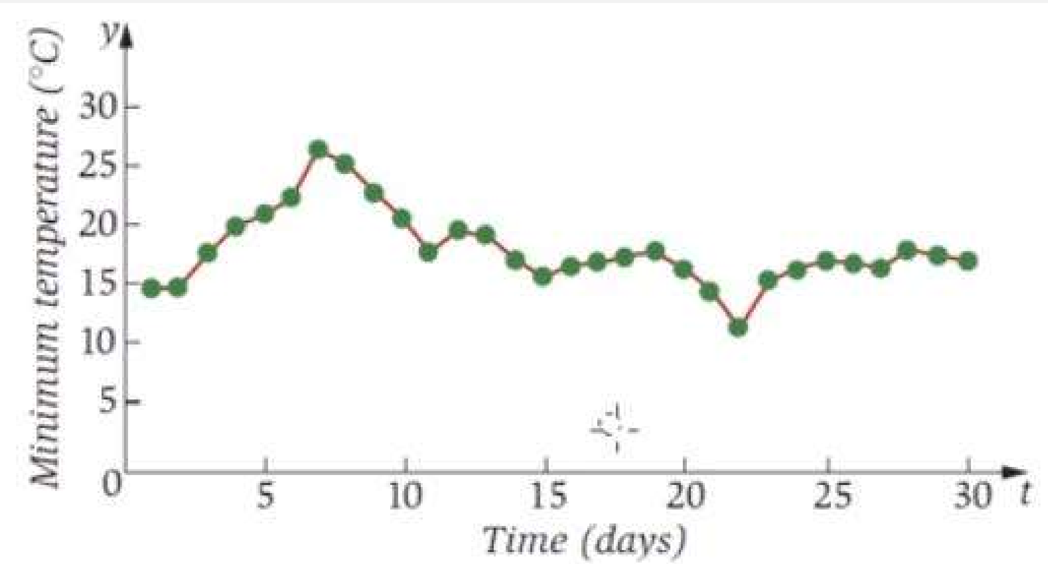
EXAMPLE

Aspect	comments
Seasonality	No
Trend	Downward secular trend
Cyclic	No
Random	Yes

EXAMPLE

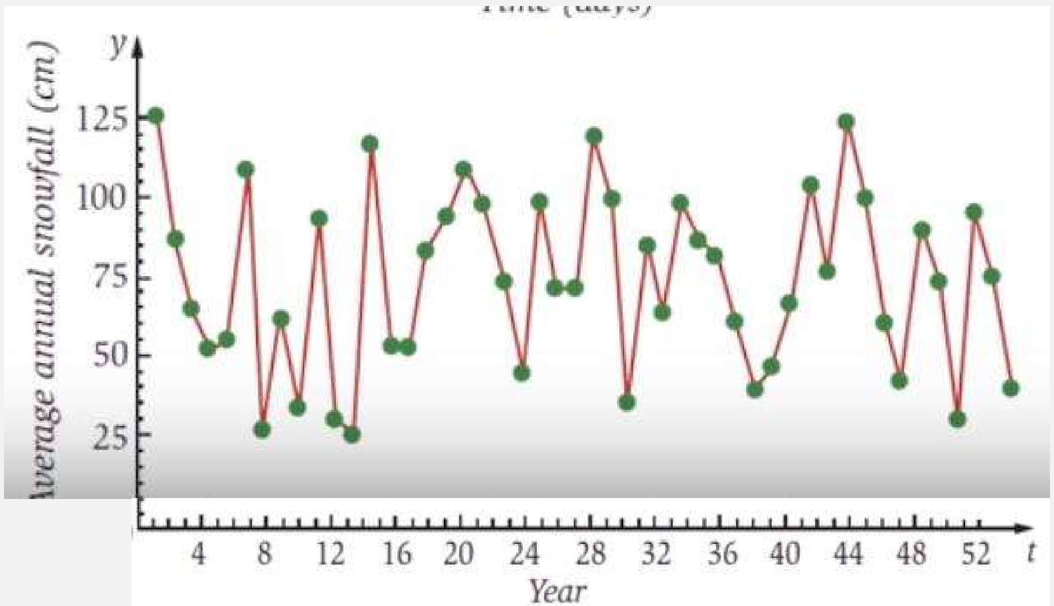
Aspect	comments
Seasonality	No
Trend	Downward secular trend
Cyclic	No
Random	Yes

EXAMPLE



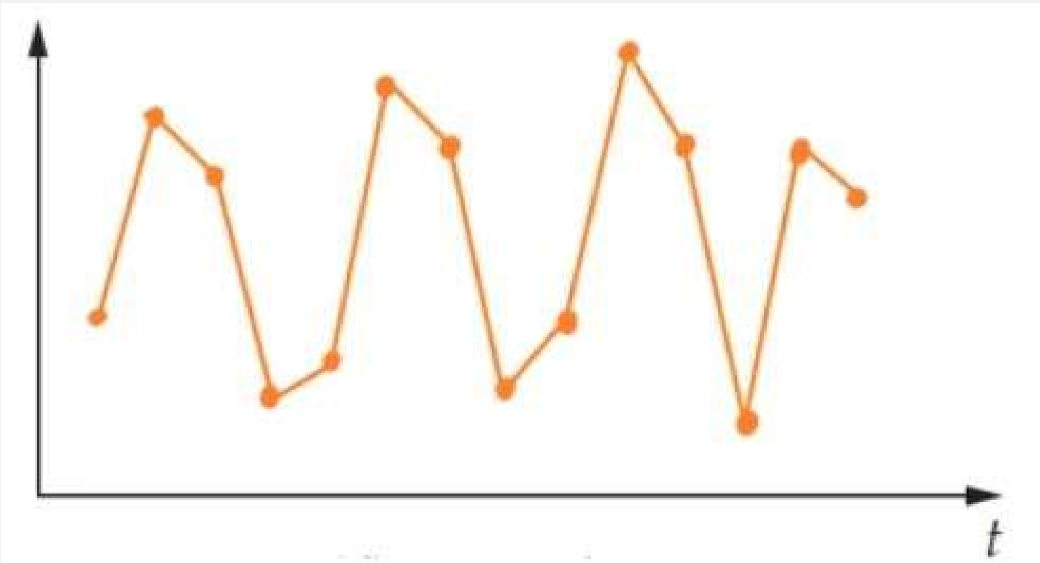
Aspect	comments
Seasonality	No
Trend	Downward secular trend
Cyclic	No
Random	Yes
Variation	Decreasing

EXAMPLE



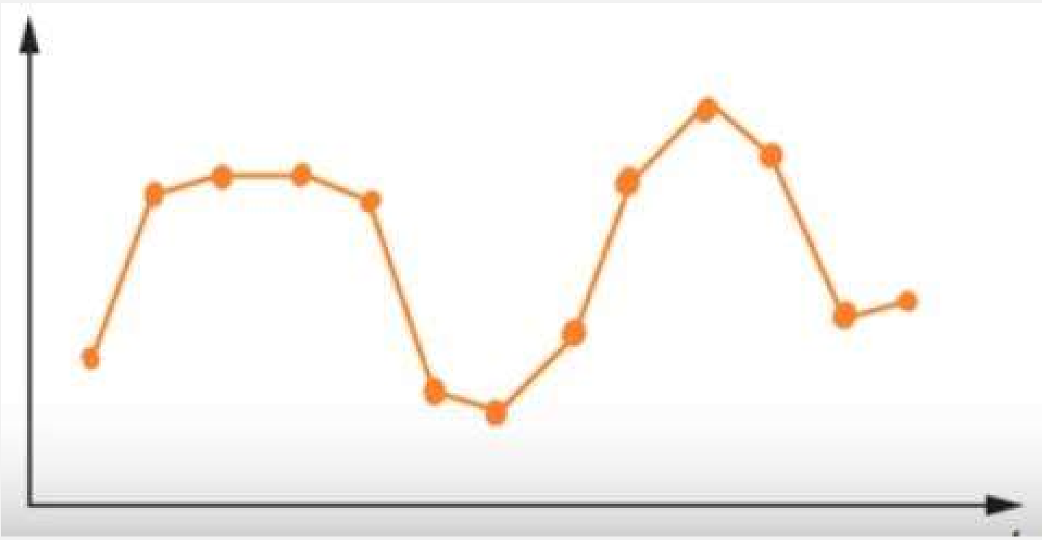
Aspect	comments
Seasonality	No
Trend	No trend
Cyclic	Yes (peaks and troughs occur at irregular points)
Random	Yes
Variation	About the same

EXAMPLE



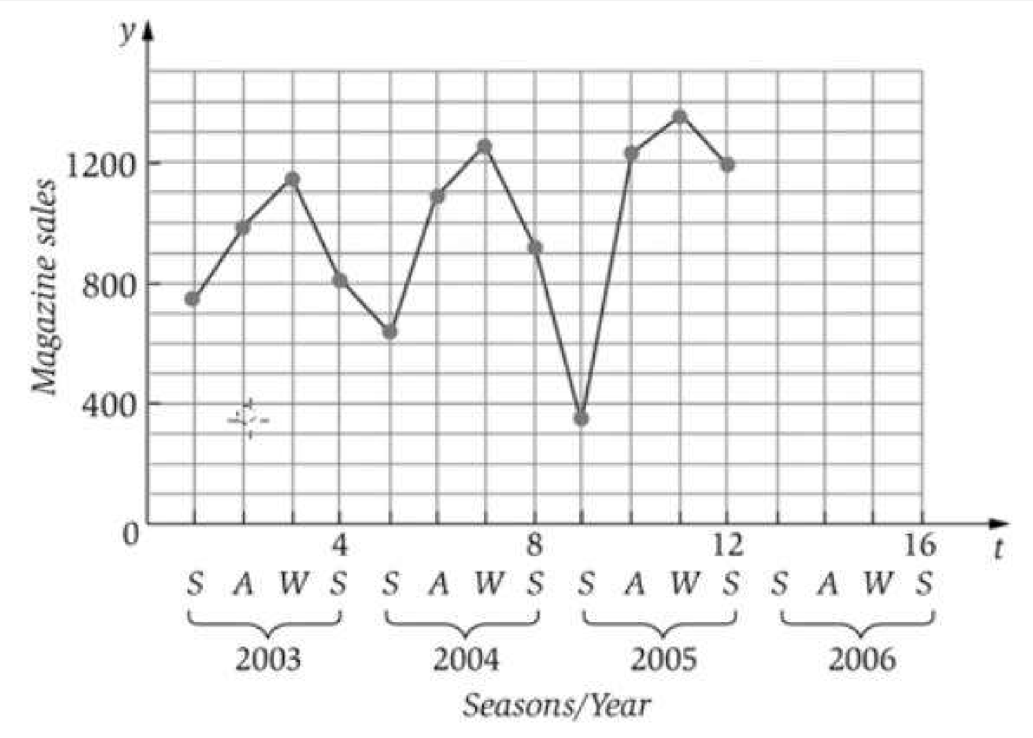
Aspect	comments
Seasonality	Yes (peaks and troughs occur at regular points)
Trend	Slight upward trend
Cyclic	No
Random	No
Variation	About the same

EXAMPLE



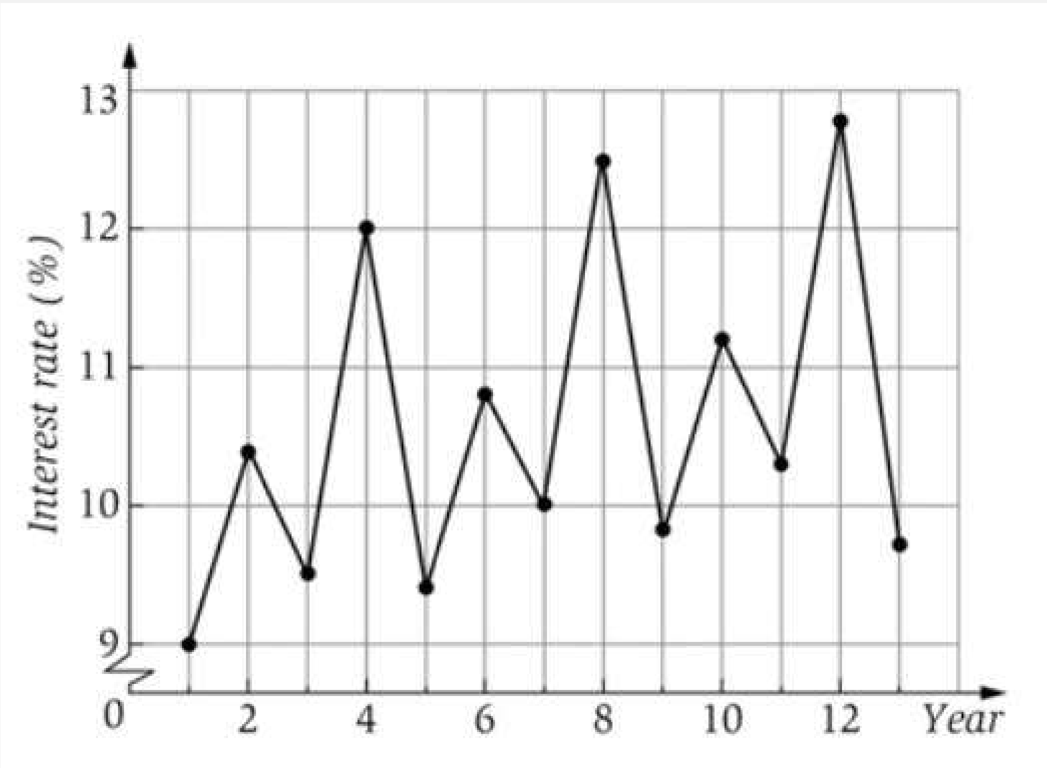
Aspect	comments
Seasonality	No (peaks and troughs occur at irregular points)
Trend	Slight upward trend
Cyclic	Yes
Random	No
Variation	About the same

EXAMPLE



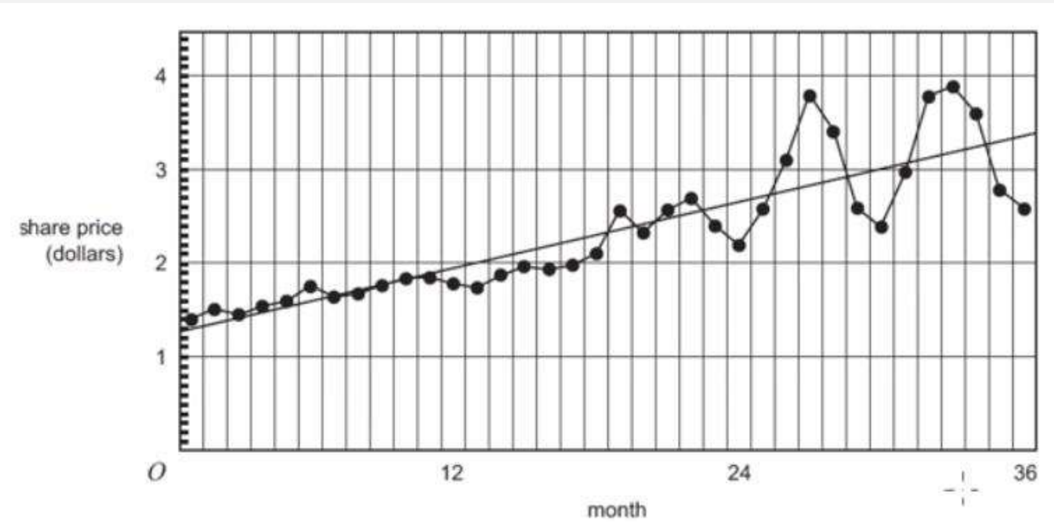
Aspect	comments
Seasonality	Yes (peaks and troughs occur at regular points)
Trend	Slight upward trend
Cyclic	No
Random	No
Variation	Increasing (cant say)

EXAMPLE



Aspect	comments
Seasonality	Yes (peaks and troughs occur at regular points)
Trend	upward trend
Cyclic	No
Random	No
Variation	About same

EXAMPLE

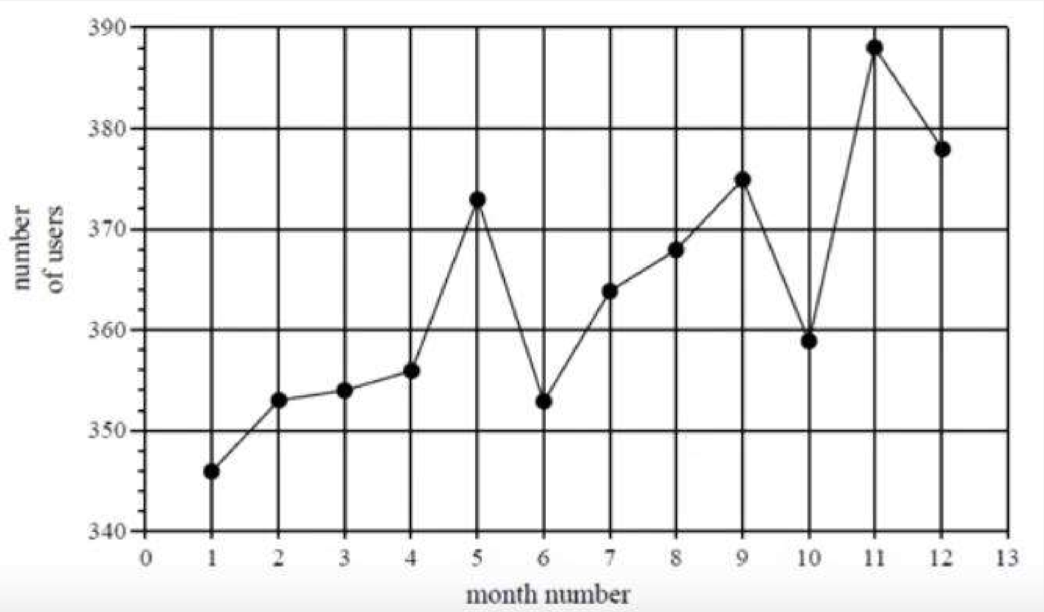


Aspect	comments
Seasonality	No
Trend	upward trend
Cyclic	No
Random	No
Variation	increasing

Which is true?

- Share price shows no trend and no change in variability?
- Share price shows no trend and increase in variability?
- Share price shows upward trend with constant variability?
- Share price shows upward trend with decreasing variability?
- Share price shows upward trend with increasing variability?

EXAMPLE

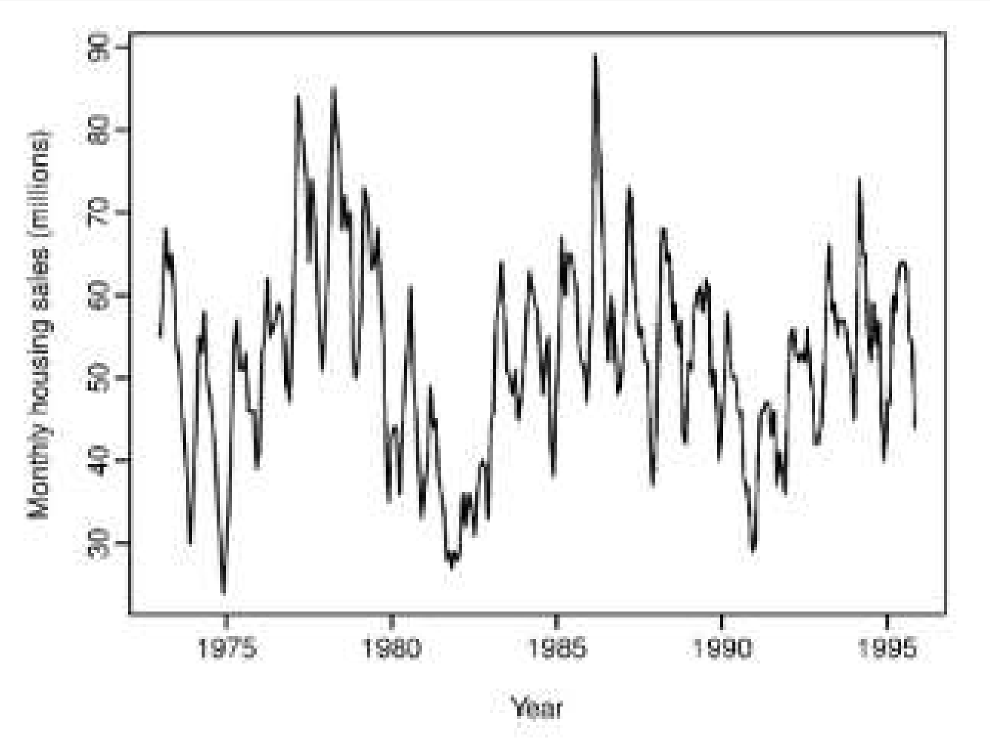


Aspect	comments
Seasonality	No
Trend	upward trend
Cyclic	No
Random	No
Variation	About the same

The time series plot has

- No trend
- No variability
- Seasonality only
- Increasing trend with seasonality
- An increasing trend only

EXAMPLES



- The monthly housing sales show strong seasonality within each year, as well as some strong cyclic behavior with period about 6–10 years.
- There is no apparent trend in the data over this period.

Aspect	comments
Seasonality	Yes
Trend	Not much
Cyclic	No
Random	No
Variation	About the same

WHITE NOISE

- A time series is white noise if the variables are independent and identically distributed with a mean of zero.
- same variance (sigma^2)
- each value has a zero correlation with all other values in the series.
- If a time series is white noise, it is a sequence of random numbers and cannot be predicted.
- If the series of forecast errors are not white noise, it suggests improvements could be made to the predictive model.

NOISE - WHY DOES IT MATTER?

2 main reasons:

Predictability: If your time series is white noise, then, by definition, it is random. We cannot reasonably model it and make predictions.

Model Diagnostics: The series of errors from a time series forecast model should ideally be white noise.

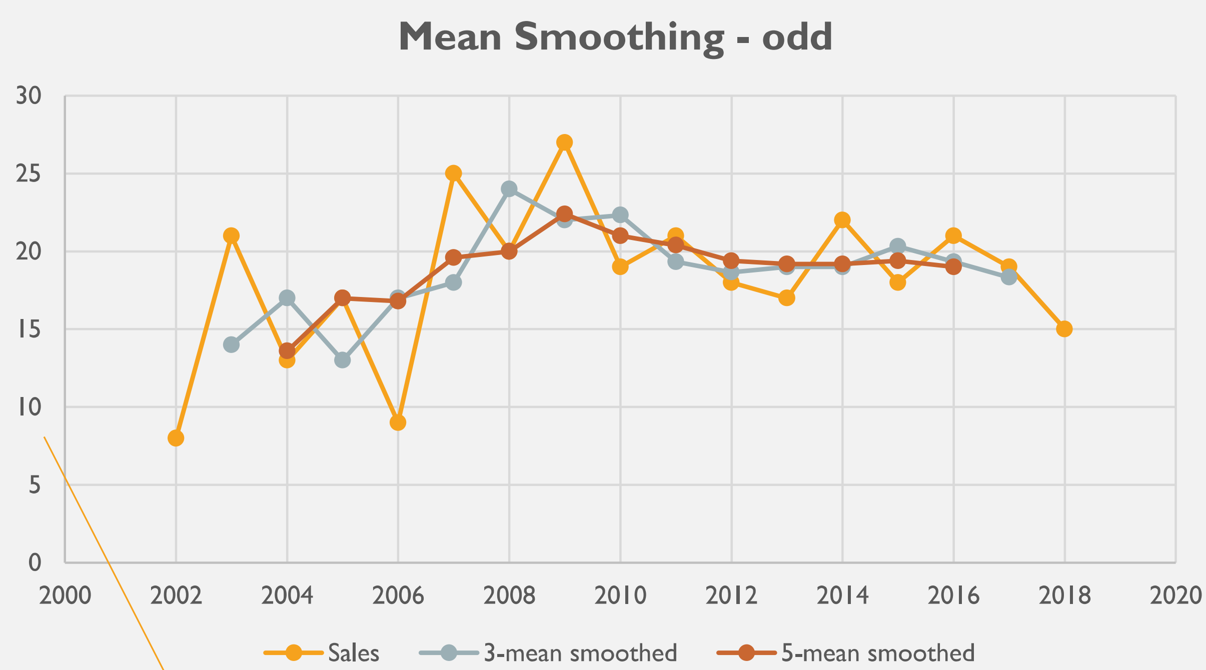
SMOOTHING

WHAT IS & WHY SMOOTHING

- Smoothing is a process of taking out the random fluctuations in the time series. Studying trend with fluctuations is erroneous
- Generally speaking, the aim of smoothing is to remove "irregular" noise and cyclical components of the time series.
- Aims to provide a more accurate prediction of the long term trend.
- Some smoothing techniques do **carry the cyclical information** forward but others would **remove** them.
- Methods to smooth
    - Using means
    - Using medians

MEAN SMOOTHING - ODD

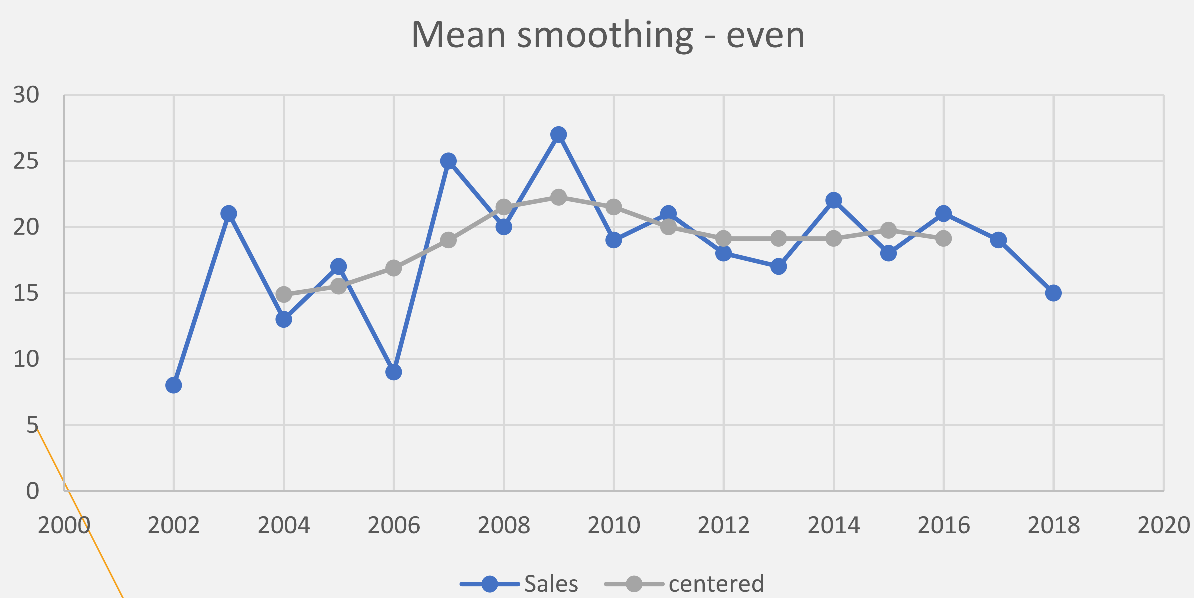
Year	Sales	3-mean smoothed	5-mean smoothed
2002	8		
2003	21	14	
2004	13	17	13.6
2005	17	13	17
2006	9	17	16.8
2007	25	18	19.6
2008	20	24	20
2009	27	22	22.4
2010	19	22.33333333	21
2011	21	19.33333333	20.4
2012	18	18.66666667	19.4
2013	17	19	19.2
2014	22	19	19.2
2015	18	20.33333333	19.4
2016	21	19.33333333	19
2017	19	18.33333333	
2018	15		



Note : be careful of over-smoothing

MEAN SMOOTHING - EVEN

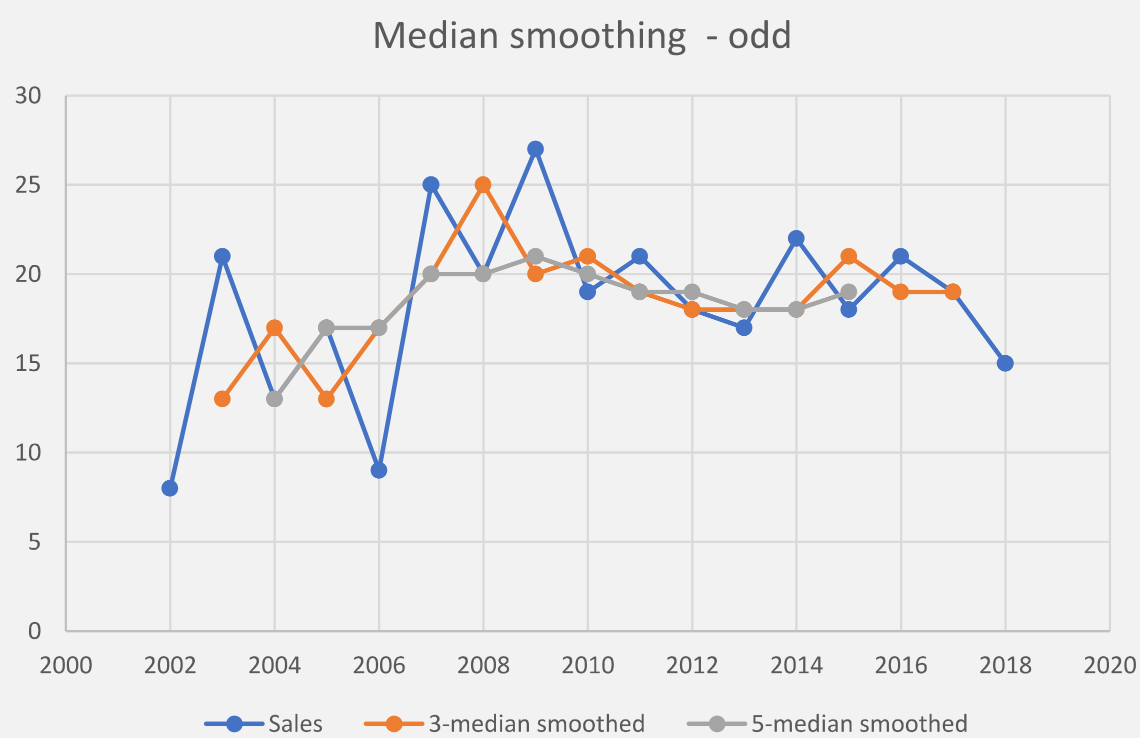
Year	Sales	4-mean smoothed	
2002	8		
2003	21	14.75	
2004	13	15	14.875
2005	17	16	15.5
2006	9	17.75	16.875
2007	25	20.25	19
2008	20	22.75	21.5
2009	27	21.75	22.25
2010	19	21.25	21.5
2011	21	18.75	20
2012	18	19.5	19.125
2013	17	18.75	19.125
2014	22	19.5	19.125
2015	18	20	19.75
2016	21	18.25	19.125
2017	19	13.75	
2018	15		



More smooth averaging, very obvious trend !!

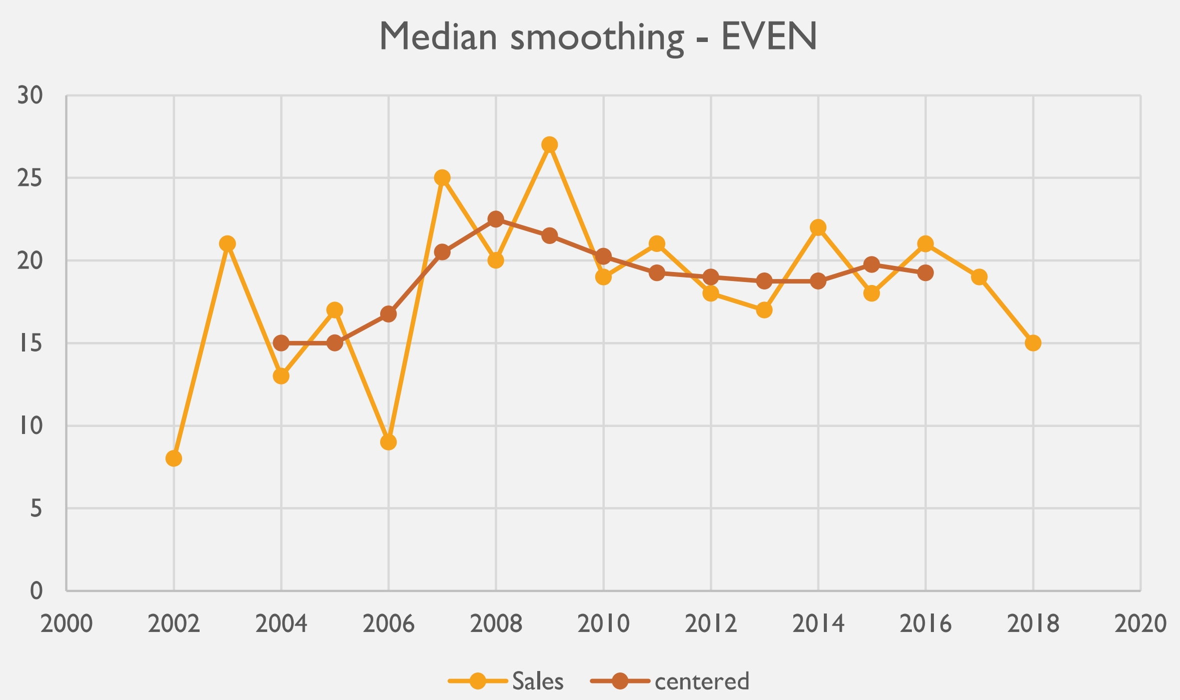
MEDIAN SMOOTHING - ODD

Year	Sales	3-median smoothed	5-median smoothed
2002	8		
2003	21	13	
2004	13	17	13
2005	17	13	17
2006	9	17	17
2007	25	20	20
2008	20	25	20
2009	27	20	21
2010	19	21	20
2011	21	19	19
2012	18	18	19
2013	17	18	18
2014	22	18	18
2015	18	21	19
2016	21	19	
2017	19	19	
2018	15		



MEDIAN SMOOTHING - ODD

Year	Sales	4-median smoothed	centered
2002	8		
2003	21	15	
2004	13	15	15
2005	17	15	15
2006	9	18.5	16.75
2007	25	22.5	20.5
2008	20	22.5	22.5
2009	27	20.5	21.5
2010	19	20	20.25
2011	21	18.5	19.25
2012	18	19.5	19
2013	17	18	18.75
2014	22	19.5	18.75
2015	18	20	19.75
2016	21	18.5	19.25
2017	19	19	
2018	15		



WHEN TO USE MEAN & MEDIAN SMOOTHING?

Mean smoothing	Median smoothing
Means are affected by outliers	Medians are not affected by outliers, generally

BASIC TIME SERIES STRUCTURES

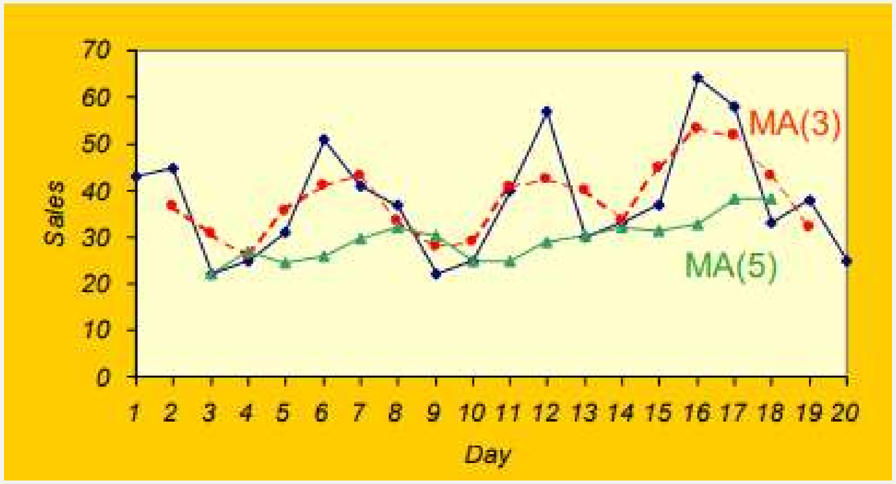
Additive Model	Multiplicative Model
$y_t = T_t + S_t + C_t + R_t$	$y_t = T_t \times S_t \times C_t \times R_t$
The additive model is useful when the seasonal variation is relatively constant over time.	The multiplicative model is useful when the seasonal variation increases over time.
different components affect the time series additively	
E.g. for monthly data, an additive model assumes that the difference between the January and July values is approximately the same each year. In other words, the amplitude of the seasonal effect is the same each year.	

METHODS FOR IDENTIFYING THE PATTERN:

SMOOTHING TECHNIQUES	Decomposition
used to remove, or at least reduce, the random fluctuations in a time series so as to more clearly expose the existence of the other components.	$y_t = T_t \times S_t \times C_t \times R_t$
2 types of smoothing techniques - Moving averages - Exponential smoothing	The multiplicative model is useful when the seasonal variation increases over time.
Moving averages for a given time period is the (arithmetic) average of the values in that time period and those close to it.	exponentially smoothed value for a given time period is the weighted average of all the available values up to that period.

SMOOTHING TECHNIQUES – MOVING AVERAGE

Day	Sales	3-day moving sum	3-day moving average
1	43		
2	45	110	36.67
3	22	92	30.67
4	25	78	26.00
5	31	107	35.67
6	51		



- No MA value for the first (neither for the last) day.
- The longer the moving average period
  - the stronger the smoothing effect,
  - the shorter the smoothed series
- When the moving average period is large, the following are removed
  - the random variations,
  - the seasonal and
  - cyclical variations
- only the long-term trend can be revealed.

Refer to python code ML-TIME-SERIES-05-smoothing

DIFFERENT CLASSICAL TIME SERIES FORECASTING METHODS

- Autoregression (AR)
- Moving Average (MA)
- Autoregressive Moving Average (ARMA)
- Autoregressive Integrated Moving Average (ARIMA)
- Seasonal Autoregressive Integrated Moving-Average (SARIMA)
- Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX)
- Vector Autoregression (VAR)
- Vector Autoregression Moving-Average (VARMA)
- Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX)
- Simple Exponential Smoothing (SES)
- Holt Winter's Exponential Smoothing (HWES)

TYPES OF MODELS

There are 2 basic types of "time domain" models.

- Models that relate the present value of a series to past values and past prediction errors - these are called ARIMA models (for Autoregressive Integrated Moving Average).

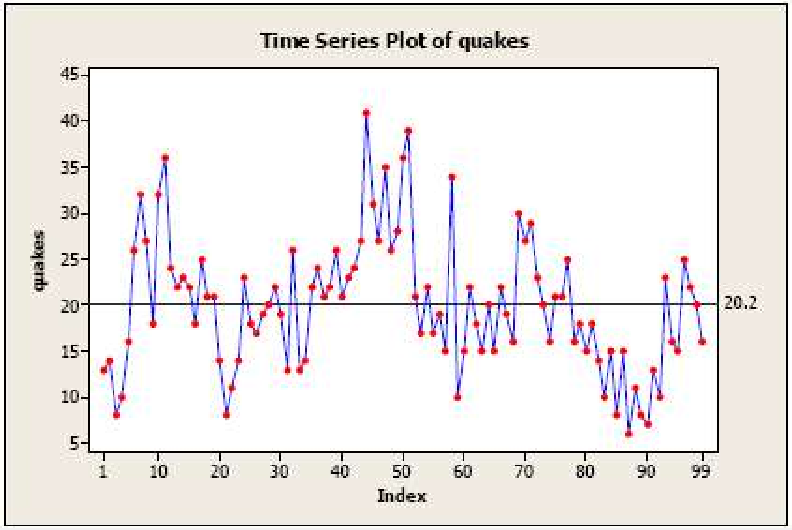
- Ordinary regression models that use time indices as x-variables. These can be helpful for an initial description of the data and form the basis of several simple forecasting methods.

IMPORTANT CHARACTERISTICS TO CONSIDER

- Is there a trend, i.e. the measurements tend to increase (or decrease) over time?
- Is there seasonality, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?
- Are their outliers? In regression, outliers are far away from the line. With time series data, the outliers are far away from the other data.
- Is there a long-run cycle or period unrelated to seasonality factors?
- Is there constant variance over time, or is the variance non-constant?
- Are there any abrupt changes to either the level of the series or the variance?

EXAMPLE

- time series plot of the annual number of earthquakes in the world with seismic magnitude over 7.0, for last 99 years.
- No consistent trend (upward or downward) over the entire time span.
- The horizontal line drawn at quakes = 20.2 indicates the mean of the series.
- Notice that the series tends to stay on the same side of the mean (above or below) for a while and then wanders to the other side.
- Almost by definition, there is no seasonality as the data are annual data.
- There are no obvious outliers.
- It's difficult to judge whether the variance is constant or not.



TIMES SERIES - MODELS

AUTOREGRESSIVE (AR)

- Autoregressive models and processes are stochastic calculations in which future values are estimated based on a weighted sum of past values.
- *model of order p*
- Autoregressive models and processes operate under the premise that past values have an effect on current values
- Regression models forecast a variable using a linear combination of predictors, whereas autoregressive models use a combination of past values of the variable.
- not applicable on non-stationary series.

AUTOREGRESSIVE (AR)

- An AR(1) autoregressive process is the first order process, meaning that the current value is based on the immediately preceding value,
- An AR(2) process has the current value based on the previous two values.
- An AR(0) process is used for white noise and has no dependence between the terms.

- Coefficients
  - many different ways to calculate the coefficients used in these calculations, including the ordinary least squares or method of movements.

  - One drawback to autoregressive models is that past prices won't always be the best predictor of future movements, if the underlying **fundamentals** of a company have changed.

MOVING AVERAGE - MA

- A moving average (MA) is a widely used indicator in technical analysis
- *model of order q*
- not applicable on non-stationary series.
- A moving average **model** is different from calculating the moving average of the time series.
- Helps smooth out price action by filtering out the "noise" from random short-term price fluctuations.

- It is a trend-following, or lagging, indicator because it is based on past prices.

- The two basic and commonly used moving averages
  - simple moving average (SMA), which is the simple average of a security over a defined number of time periods,
  - exponential moving average (EMA), which gives greater weight to more recent prices.

MOVING AVERAGE - MA – WHAT IT TELLS

- MA lag current price action because they are based on past prices;
  - The longer the time period for the moving average, the greater the lag.
  - 200-day MA will have a much greater degree of lag than a 20-day MA

- The length of the moving average to use depends on the trading objectives,
  - with shorter moving averages used for short-term trading and longer-term moving averages more suited for long-term investors.
  - The 50-day and 200-day MAs are widely followed by investors and traders, with breaks above and below this moving average considered to be important trading signals.

- A rising moving average indicates that the security is in an uptrend, while a declining moving average indicates that it is in a downtrend.

MOVING AVERAGE - MA

- The most common applications of moving averages are to
  - identify the trend direction, and
  - to determine support and resistance levels.

AUTOREGRESSIVE MOVING AVERAGE – (ARMA)

- model of order p, q

ARIMA

- generalization of an autoregressive moving average (ARMA) model
- An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either
  - better understand the data set or
  - to predict future trends.

- An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

BREAKING DOWN - ARIMA

- **components**

  - Autoregression (AR) refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

  - Integrated (I) represents the differencing of raw observations to allow for the time series to become stationary, i.e., data values are replaced by the difference between the data values and the previous values.

  - Moving average (MA) incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

ARIMA - NOTATIONS

- ARIMA models are typically expressed like "ARIMA(p, d, q)",
- ARIMA - standard notation
  - p, d, and q
    - where integer values substitute for the parameters to indicate the type of ARIMA model used.
  - The parameters can be defined as:
    - p: the number of lag observations in the model; also known as the lag order. This captures the "autoregressive" nature of ARIMA.
    - d: the number of times that the raw observations are differenced; also known as the degree of differencing. This captures the "integrated" nature of ARIMA
    - q: the size of the moving average window; also known as the order of the moving average.

USING TS MODELS

Criteria	Choice	Comments
data does not have a trend or a seasonal component	Moving Average Single Exp Smoothing	
data has a trend but does not have a seasonal component	Trend Analysis Double Exp Smoothing	Trend Analysis fits a single equation to the data, which works well when the trend follows a consistent shape without shifts or reversals. Double Exponential Smoothing uses a dynamic trend component that works well when the data have cyclical movements, shifts in the trend, or even reversals in the trend.
data have a seasonal component	Decomposition Winters' Method	Use Winters' Method when you want to use your time series model to generate forecasts. Usually, you should not use Decomposition to generate forecasts, but it can be useful to examine the components of the time series. For example, you could use Decomposition to communicate time series concepts to management.

QUESTIONS

QUESTION

1) **Which of the following is an example of time series problem?**

   1. **Estimating number of hotel rooms booking in next 6 months.**
   2. **Estimating the total sales in next 3 years of an insurance company.**
   3. **Estimating the number of calls for the next one week.**

A) Only 3
B) 1 and 2
C) 2 and 3
D) 1 and 3
E) 1,2 and 3

QUESTION

2) Which of the following is not an example of a time series model?

A) Naive approach
B) Exponential smoothing
C) Moving Average
D) None of the above

- 3) Which of the following can’t be a component for a time series plot?

- A) Seasonality
- B) Trend
- C) Cyclical
- D) Noise
- E) None of the above

- Solution: (E)