Slides Prepared By
Dr. Hussein Hazimeh

Lebanese University

Faculty of Information 1

Data Science Departement

2nd year – Data Analysis in R

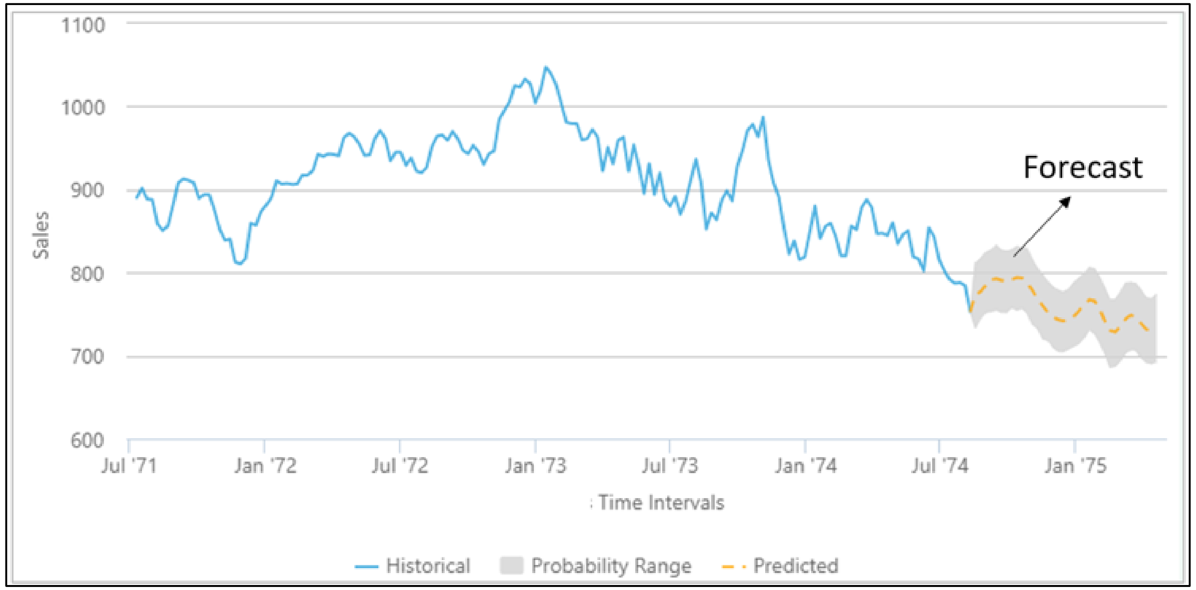
May – 2022

Agenda

- Time series analysis
- Moving average models
- Winters model
- Exponential smoothing techniques
- Regression models
- Measures for godness of the forecast
- Handling missing values in time series

Time series analysis

» Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals.

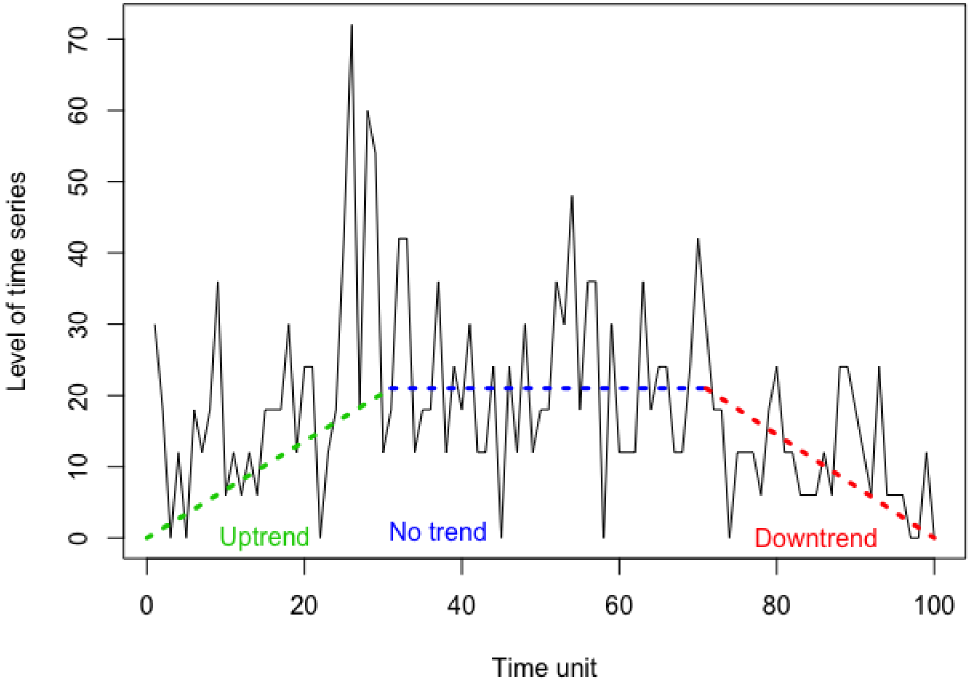


» Time series components:

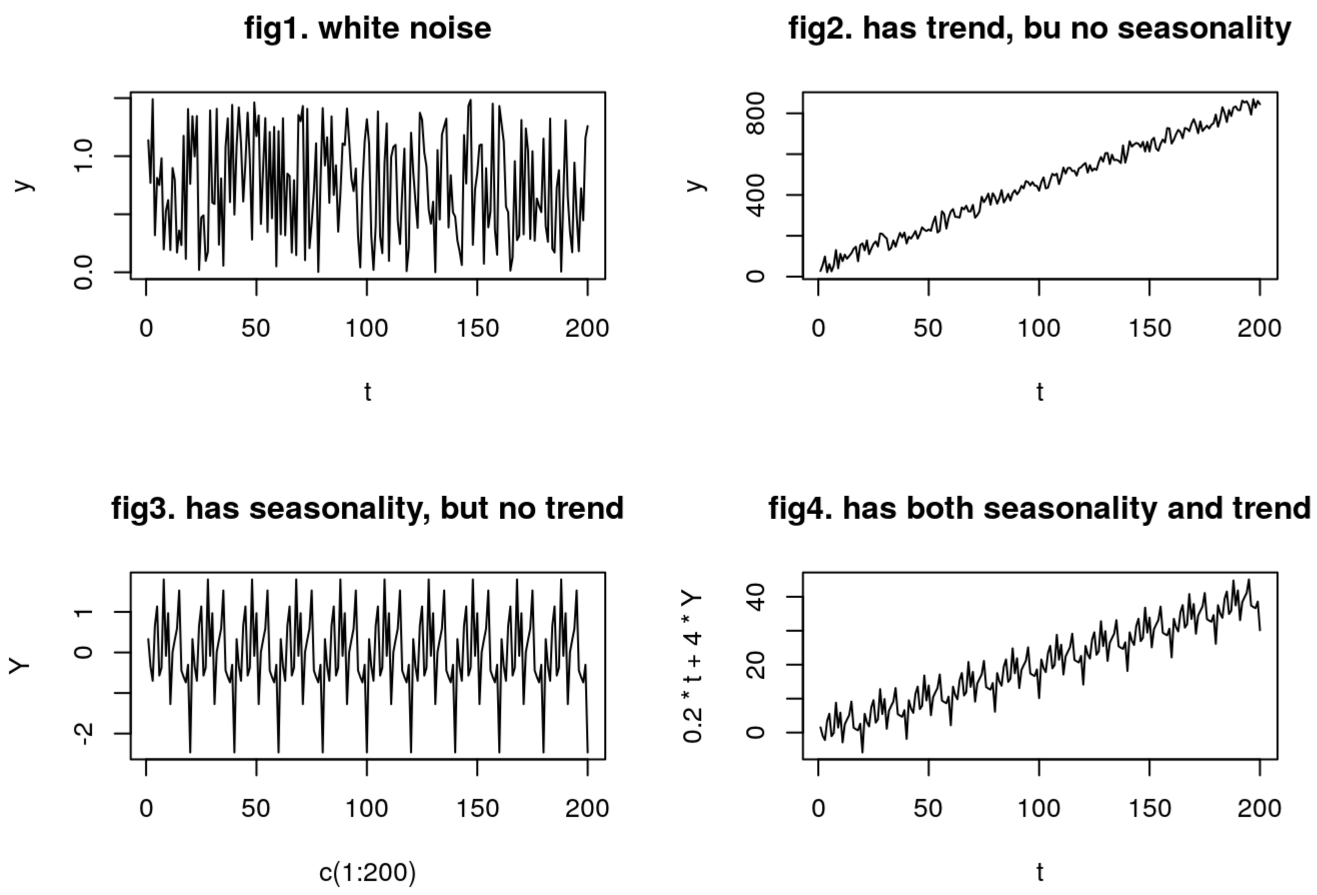
- Trend
- Seasonality

Time series analysis (Trend)

- A trend is a consistent increase or decrease in units of the underlying data.
- Also, if there is no trend, this can be called stationary.
-



Types of time series



Moving average

A **moving average** (**rolling average** or **running average**) is a calculation to analyze data points by creating a series of averages of different subsets of thes full data set.

A **moving average** can smooth out the noise of random outliers and emphasize long-term trends

Moving average is composed of 3 types:

1. Simple Moving Average (SMA)
2. Weighted Moving Average (WMA)
3. Exponential Moving Average (EMA)

Simple Moving Average

Year	Sales (GWh)	5-MA
1989	2354.34	
1990	2379.71	
1991	2318.52	2381.53
1992	2468.99	2424.56
1993	2386.09	2463.76
1994	2569.47	2552.60
1995	2575.72	2627.70
1996	2762.72	2750.62
1997	2844.50	2858.35
1998	3000.70	3014.70
1999	3108.10	3077.30
2000	3357.50	3144.52
2001	3075.70	3188.70
2002	3180.60	3202.32
2003	3221.60	3216.94
2004	3176.20	3307.30
2005	3430.60	3398.75
2006	3527.48	3485.43
2007	3637.89	
2008	3655.00	

$$MA_n = \frac{\sum_{i=1}^{n} D_i}{n}$$

Where,

n = number of periods in the moving average

D_i = demand in period i

Simple Moving Average

Date	Closing Price of AAPL
June 26	$90.90
June 25	$90.36
June 24	$90.28
June 23	$90.83
June 20	$90.91

A five-period moving average, based on the prices above, would be calculated using the following formula:

$$\text{MA} = \frac{P_1 + P_2 + P_3 + P_4 + P_5}{5}$$

where:

P_n = Price for time period

or:

$$\frac{90.90 + 90.36 + 90.28 + 90.83 + 90.91}{5} = 90.656$$

Weighted Moving Average

Date	Closing Price of AAPL	Weighting
June 26	$90.90	5/15
June 25	$90.36	4/15
June 24	$90.28	3/15
June 23	$90.83	2/15
June 20	$90.91	1/15

The weighted average is calculated by multiplying the given price by its associated weighting and totaling the values. The formula for the WMA is as follows:

$$\text{WMA} = \frac{\text{Price}_1 \times n + \text{Price}_2 \times (n-1) + \cdots \text{Price}_n}{\frac{n \times (n+1)}{2}}$$

where:

$n = \text{Time period}$

The denominator of the WMA is the sum of the number of price periods as a triangular number. In the example from the table above, the weighted five-day moving average would be $90.62:

$$(90.90 \times \tfrac{5}{15}) + (90.36 \times \tfrac{4}{15}) + (90.28 \times \tfrac{3}{15}) + (90.83 \times \tfrac{2}{15}) + (90.91 \times \tfrac{1}{15}) = \$90.62$$

Simple Moving Average: example

Year	Sales (GWh)	5-MA
1989	2354.34	
1990	2379.71	
1991	2318.52	2381.53
1992	2468.99	2424.56
1993	2386.09	2463.76
1994	2569.47	2552.60
1995	2575.72	2627.70
1996	2762.72	2750.62
1997	2844.50	2858.35
1998	3000.70	3014.70
1999	3108.10	3077.30
2000	3357.50	3144.52
2001	3075.70	3188.70
2002	3180.60	3202.32
2003	3221.60	3216.94
2004	3176.20	3307.30
2005	3430.60	3398.75
2006	3527.48	3485.43
2007	3637.89	
2008	3655.00	

2354.34+2379.71+2381.52+2468.99+2386.09 / 5 = 2381.53

$$MA_n = \frac{\sum_{i=1}^{n} D_i}{n}$$

Where,

n = number of periods in the moving average

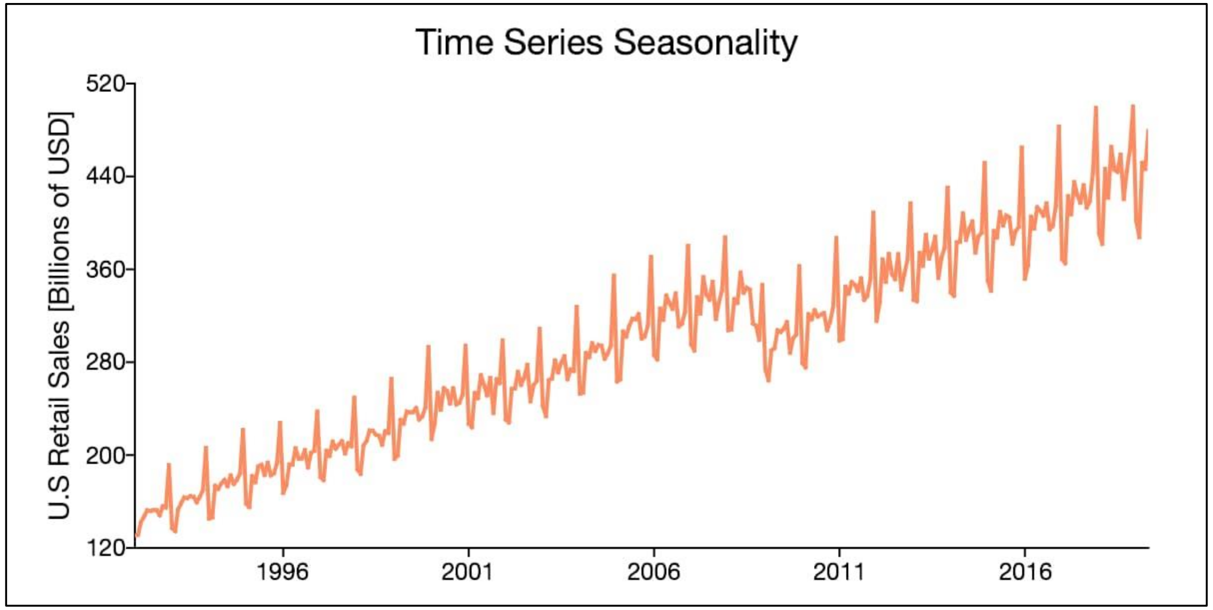
D_i = demand in period i

Seasonality Models

Seasonality models

» Seasonality is another characteristic of time series data that can be visually identified in time series plots. Seasonality occurs when time series data exhibits regular and predictable patterns at time intervals that are smaller than a year.

» An example of a time series with seasonality is flight tickets sales, which often increase between June to September and will decrease between January to May.



Winters model (seasonality)

Case study: demands per year

Demand Matrix

	Year 1	Year 2	Year 3
Q1	53	58	62
Q2	22	25	27
Q3	37	40	44
Q4	45	50	56
Total	157	173	189

+16 slope

Year 4

.... ? F_{13}

→ Level values

$$a_{t+1} = \propto \left(\frac{D_{t+1}}{C_{t+1}}\right) + (1-\propto)(a_t + b_t)$$ Level

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

$$C_{t+p+1} = \gamma \left(\frac{D_{t+1}}{a_{t+1}}\right) + (1 + \gamma)\, C_{t+1}$$ Seasonality

$$F_{t+1} = (a_t + b_t)\, C_{t+1}$$

Winters model (seasonality)

Seasonality ($C1$) = $\frac{D_t}{Level} = \frac{53}{157} = 0.34$

	Year 1	Year 2	Year 3
Q1	53	58	62
Q2	22	25	27
Q3	37	40	44
Q4	45	50	56
Total	157	173	189

Seasonaity Matrix

	Year 1	Year 2	Year 3
Q1	0.34	0.34	0.32
Q2	0.14	0.14	0.14
Q3	0.24	0.24	0.23
Q4	0.29	0.29	0.30

Winters model (seasonality)

Level (a_1) = $\frac{D_t}{C_1} = \frac{53}{0.34} = 156$

Level (a_2) = $a_{t+1} = \propto \left(\frac{D_{t+1}}{C_{t+1}}\right) + (1 - \propto)(a_t + b_t)$

= 0.2 (156) + 0.8 (156 + 4) = 159.43

	Year 1	Year 2	Year 3
Q1	53	58	62
Q2	22	25	27
Q3	37	40	44
Q4	45	50	56
Total	157	173	189

Level Matrix

	Year 1	Year 2	Year 3
Q1	156	…	…
Q2	159.43	…	…
Q3	…	…	…
Q4	…	…	…

b1 = 16 / 4 = 4

Winters model (seasonality)

Trend (b_2) = $b_{t+1} = \beta(a_{t+1} - a_t) + (1-\beta)\, b_t$

= 0.3 (159.43-156) + 0.7 (4) = 3.829

	Year 1	Year 2	Year 3
Q1	53	58	62
Q2	22	25	27
Q3	37	40	44
Q4	45	50	56
Total	157	173	189

Trend Matrix

	Year 1	Year 2	Year 3
Q1	4	…	…
Q2	3.82	…	…
Q3	…	…	…
Q4	…	…	…

b1 = 16 / 4 = 4

Winters model (seasonality)

Future Value (F_3) = $F_{t+1} = (a_t + b_t)\ C_{t+1}$

= (159.43 + 3.82) 0.24

= 39.18

	Year 1	Year 2	Year 3
Q1	53	58	62
Q2	22	25	27
Q3	37	40	44
Q4	45	50	56
Total	157	173	189

New Demand Matrix (Future Values)

	Year 1	Year 2	Year 3	Year 4
Q1	53	…	…	67.29
Q2	22.4	…	…	
Q3	39.18	…	…	
Q4	…	…	…	

Measures for godness of forecast

- Mean Absolute Deviation (MAD)
- Mean Squared Deviation (MSD)
- Mean % Deviation (M%D)

Mean Absoulte Deviation

» **Mean Absolute Deviation (MAD)**

$$\sum_{i=1}^{n} |L_t - MA(x)|$$

» Consider the following level data {30,32,31,30} where the **SMA** = 30.75 or F5 = 30.75

MAD = |30-30.75|+|32-30.75|+|31-30.75|+|30-30.75|

= 0.75 + 1.25 +0.25 +0.75

= 3

Mean Squared Deviation

- **Mean Squared Deviation (MSD)**

$$\sum_{i=1}^{n}(L_t - MA(x))^2$$

- Consider the following level data {30,32,31,30} where the **SMA** = 30.75 or F5 = 30.75

MSD = $(30-30.75)^2+(32-30.75)^2+(31-30.75)^2+(30-30.75)^2$

= 2.75

Mean %Deviation

» **Mean % Deviation (M%D)**

$$\sum_{i=1}^{n}\left(\frac{|L_t - MA(x)|}{MA(x)}\right) \times 100$$

» Consider the following level data {30,32,31,30} where the **SMA** = 30.75 or F5 = 30.75

M%D = $\frac{0.75}{30.75} \times 100 + \frac{1.25}{30.75} \times 100 + \frac{0.25}{30.75} \times 100 + \frac{0.75}{30.75} \times 100$

Finding missing value(s) in a time series

Strategy			
Linear interpolation	average method	Max-min	Actual value
41.3	44	40	40

	Year 1	Year 2	Year 3
Q1	53	58	62
Q2	22	25	27
Q3	37	?	44
Q4	45	50	56
Total	157	173	189

$$x_{t-1} + \frac{x_{t+1} - x_{t-1}}{y_{t+1} - y_{t-1}} \times (y_t - y_{t-1})$$

References

» https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm

» https://otexts.com/fpp2/holt-winters.html