

Chat

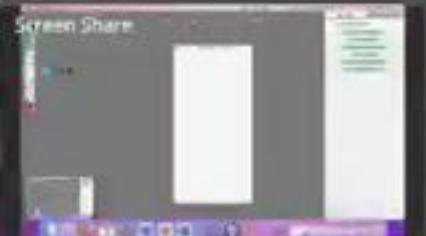
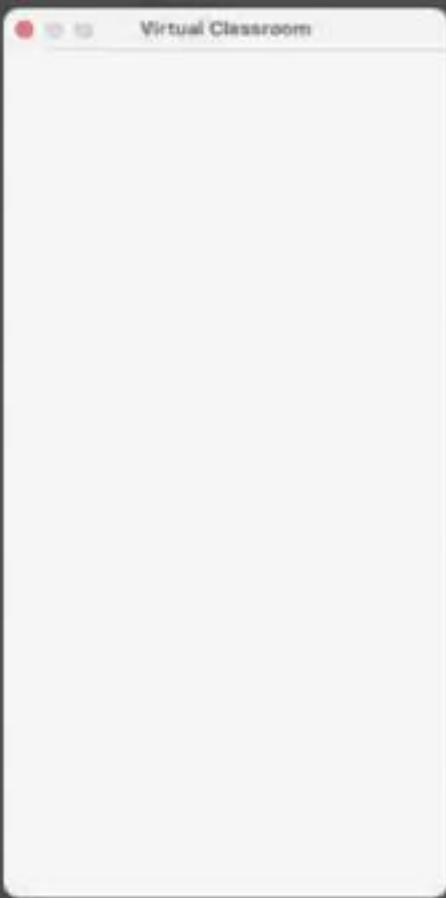
Participants (13)

 Disable group chat

NEKODIA DIVASTAVI joined

DINESH BHAGATI joined





Everyone  
Write message >



## Chat

## Participants (17)

 Disable group chat

MUGDHA SRIVASTAVA joined

DIRE BHASKAR joined

PEDDADA CHAKRABARTHY joined

RAJESWARI joined

NAGAVI GANESH NAYAK joined

SHYAMALA DOWRVIGO joined

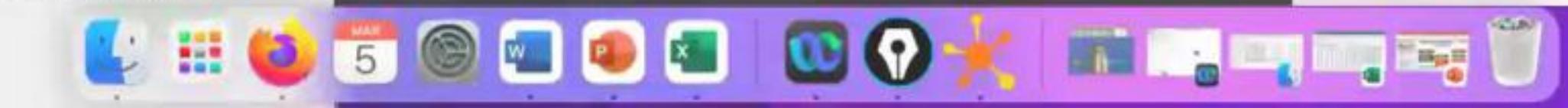


Screen Share

Virtual Classroom

Chat Participants (18)

Disable group chat



Chat Participants (17)

Disable group chat

MSASDA SRIVASTAVA joined

DIRE BHASKAR joined

PEDDADA CHAKRavarthy joined

RAJESH A joined

NAGARAJ GANESH NAYAK joined

SHYAMALA DOMBIVI joined

HARSH VADSHIRI joined

Statistical Methods for Data Science

## Time Series ➔ Types of forecasting methods

## Forecasting Methods

### **Qualitative**

- Personal Opinion or Judgment
  - Panel consensus
  - Delphi Method
  - Market Research

## Quantitative

Time Series

- Smoothing method
  - Exponential smoothing
  - Trend projection method

Causa

- ## Regression Trend Analysis

Statistical Methods for Data Science

## Time Series → Types of forecasting methods



• Chat

 Participants (23)

Disable group chat

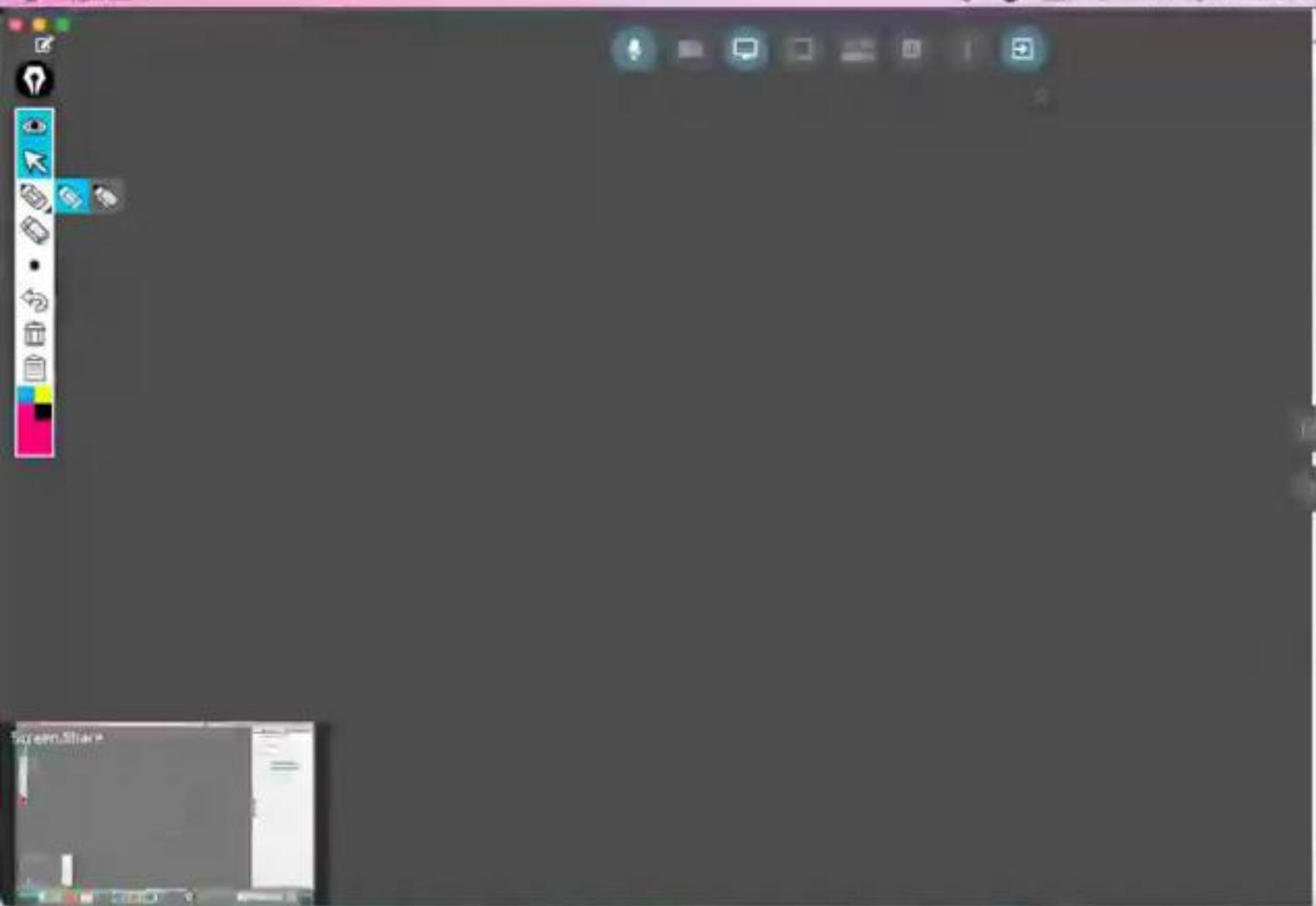
• 1992 年 10 月刊第 1 号

PRIMERAS REACCIONES

www.scholarship.org

aboraiah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore



## Chat

## Participants (30)

 Disable group chat

Gurugobalsan

Great evening Sir

Rakesh Rajput

good evening Sir

SAKSHAM JAIN

HARSHAD SHRIKANT JAIN

## Time Series → Qualitative forecasting



- **Personal Opinion**

- Individuals forecasts future based on their own judgment or opinion without any formal model
- Such assessments are relatively reliable and accurate

## Time Series → Qualitative forecasting

### Delphi Method

The aim of the Delphi method is to construct consensus forecasts from a group of experts in a structured iterative manner. A facilitator is appointed in order to implement and manage the process. The Delphi method generally involves the following stages:

## Time Series → Qualitative forecasting

### Delphi Method

- A panel of experts is assembled.
- Forecasting tasks/challenges are set and distributed to the experts.
- Experts return initial forecasts and justifications. These are compiled and summarised in order to provide feedback.

Virtual Classroom



Participants (36)

Disable group chat

G

aboraiah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

## Delphi Method

- Feedback is provided to the experts, who now review their forecasts in light of the feedback. This step may be iterated until a satisfactory level of consensus is reached.
- Final forecasts are constructed by aggregating the experts' forecasts.

# Statistical Methods for Data Science

Time Series → Qualitative forecasting



Virtual Classroom

Chat

Participants (36)

Disable group chat

ethod

Abdullah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore

## Time Series → Qualitative forecasting

### Delphi Method

- Each stage of the Delphi method comes with its own challenges. In what follows, we provide some suggestions and discussions about each one of these.

## Time Series → Definition

A time series is defined as a set of observations on a variable generated sequentially in time. The measurement of the variable may be made continuously or at discrete (equally spaced) intervals.

Often a variable continuous in time is measured at regular intervals and this produces a discrete series of data.

## Time Series → Definition

Thus a time series may be represented as a set of data ( $y_1, y_2, y_3, \dots, y_n$ ),  $y_t$  denoting the value of the variable  $y$  at time  $t$ .

Example of continuous variable:

- Temperature on a chemical reactor
- Level of tide at a particular site
- Amplitude of an electrical signal

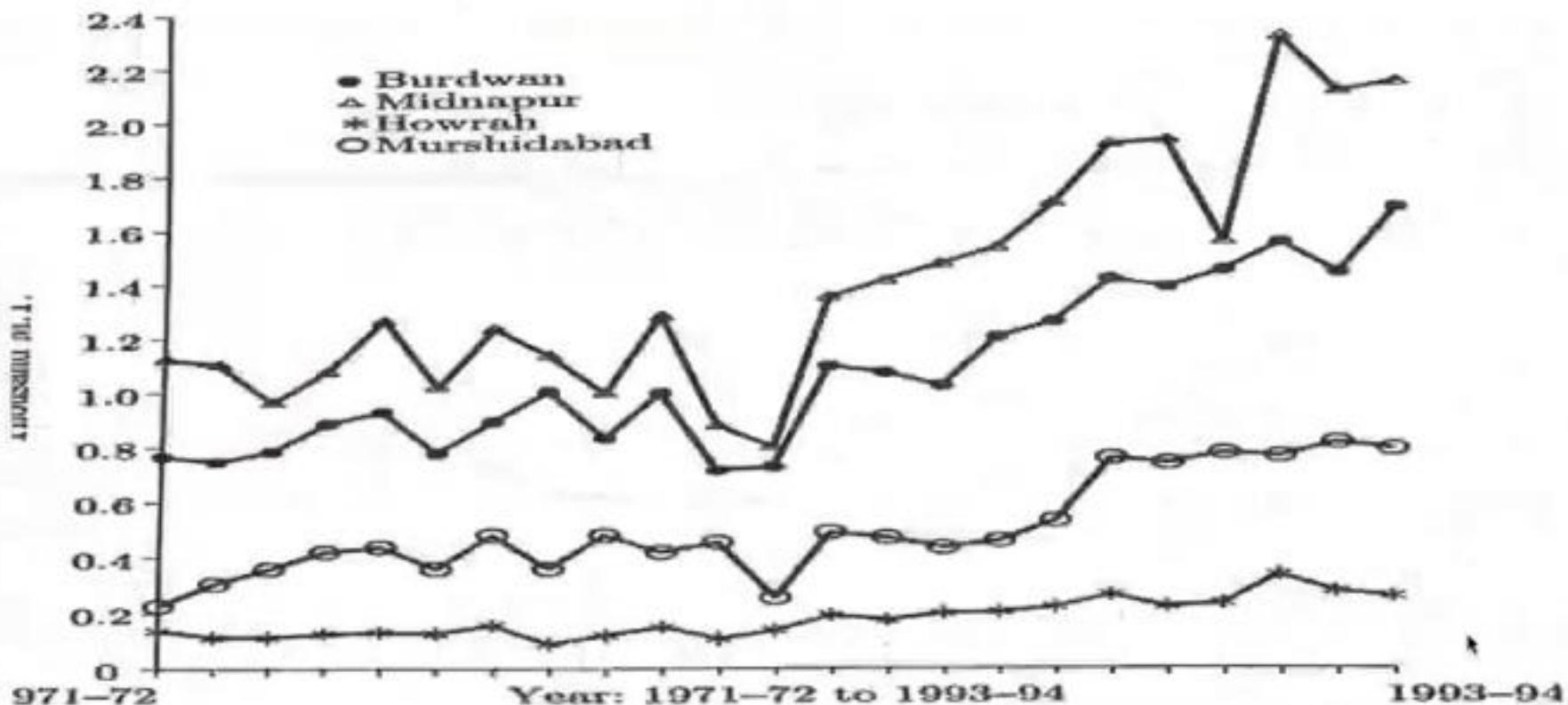


Fig 14.2 District-wise production of rice for four districts in West Bengal.

Virtual Classroom

Chat

Participants (43)

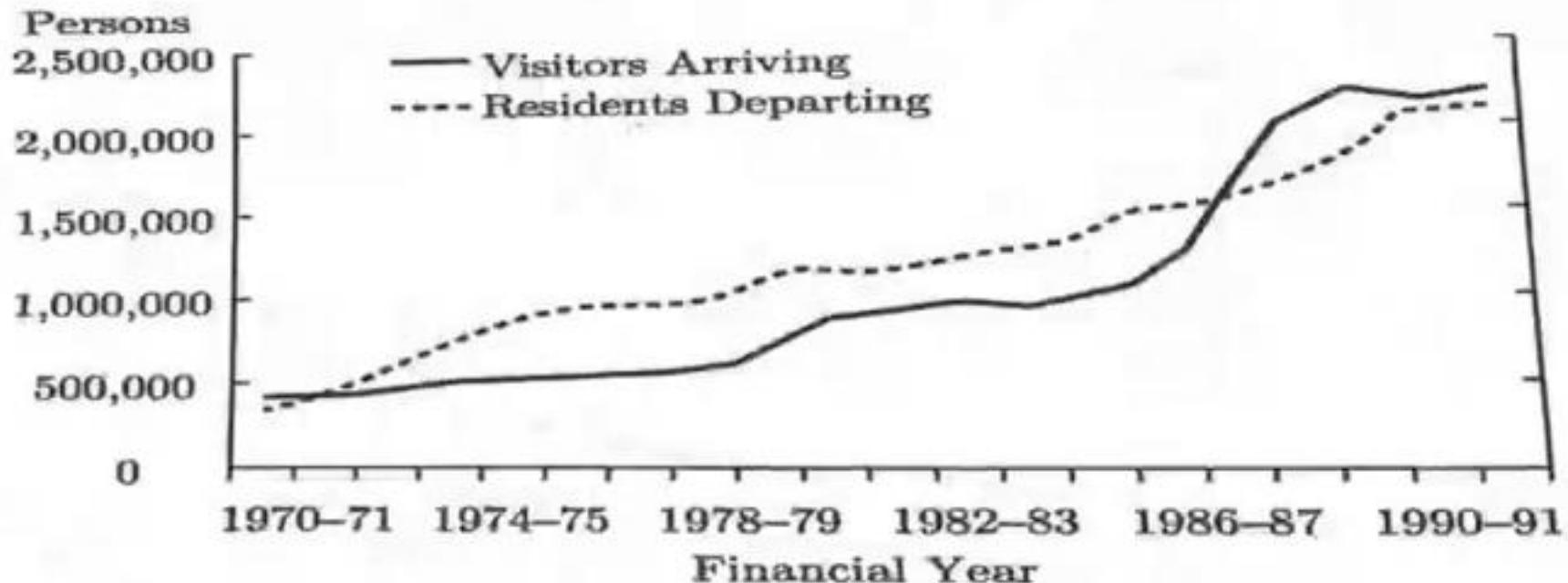
Disable group chat

PRIYAMOHAN JOINED

# Methods for Data Science

Series ➔

Definition



**Fig 14.3** Short-term movement to and from Australia.

[Source: Immigration update, December, Quarter 1991, Bureau of Immigration Research, Statistics Section, Australia.]

Virtual Classroom

Chat

Participants (43)

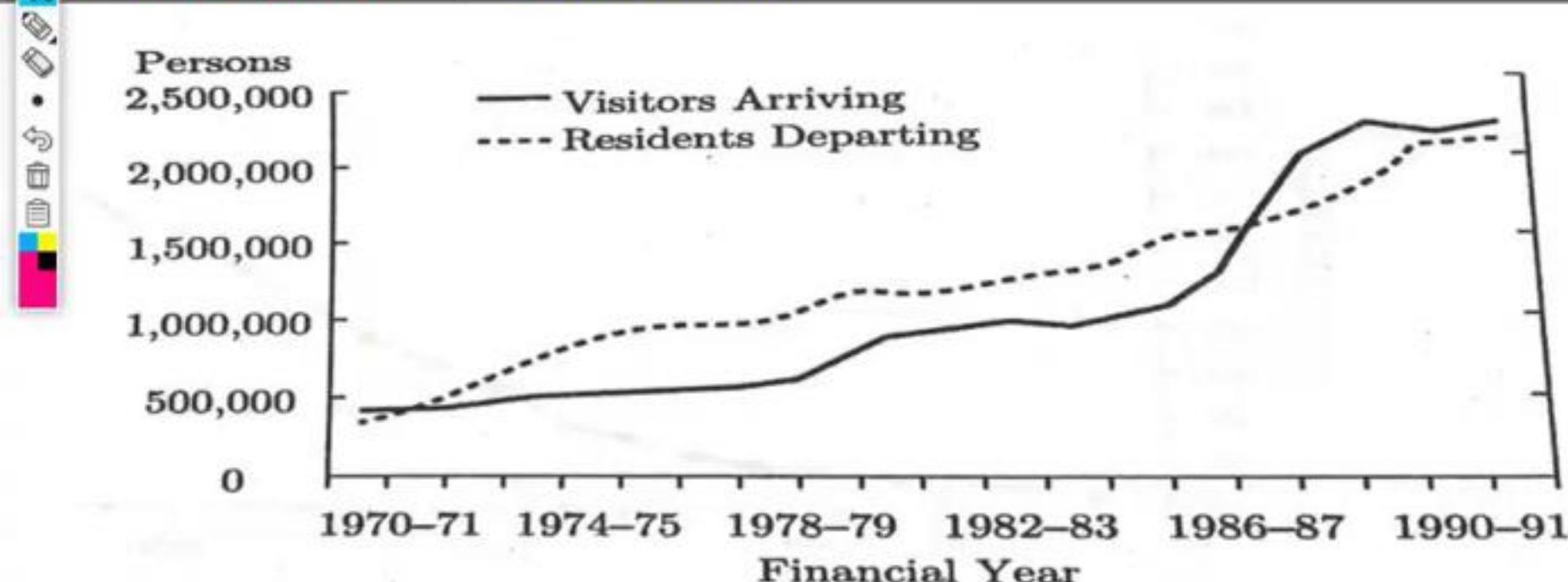
Disable group chat

aboraiah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

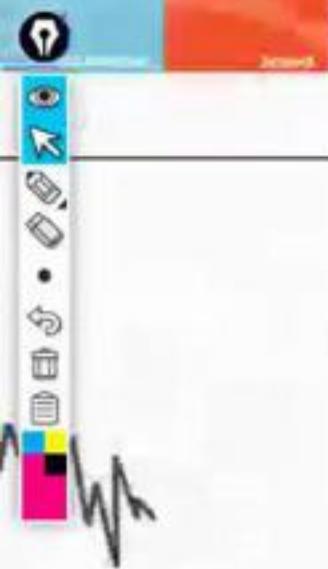
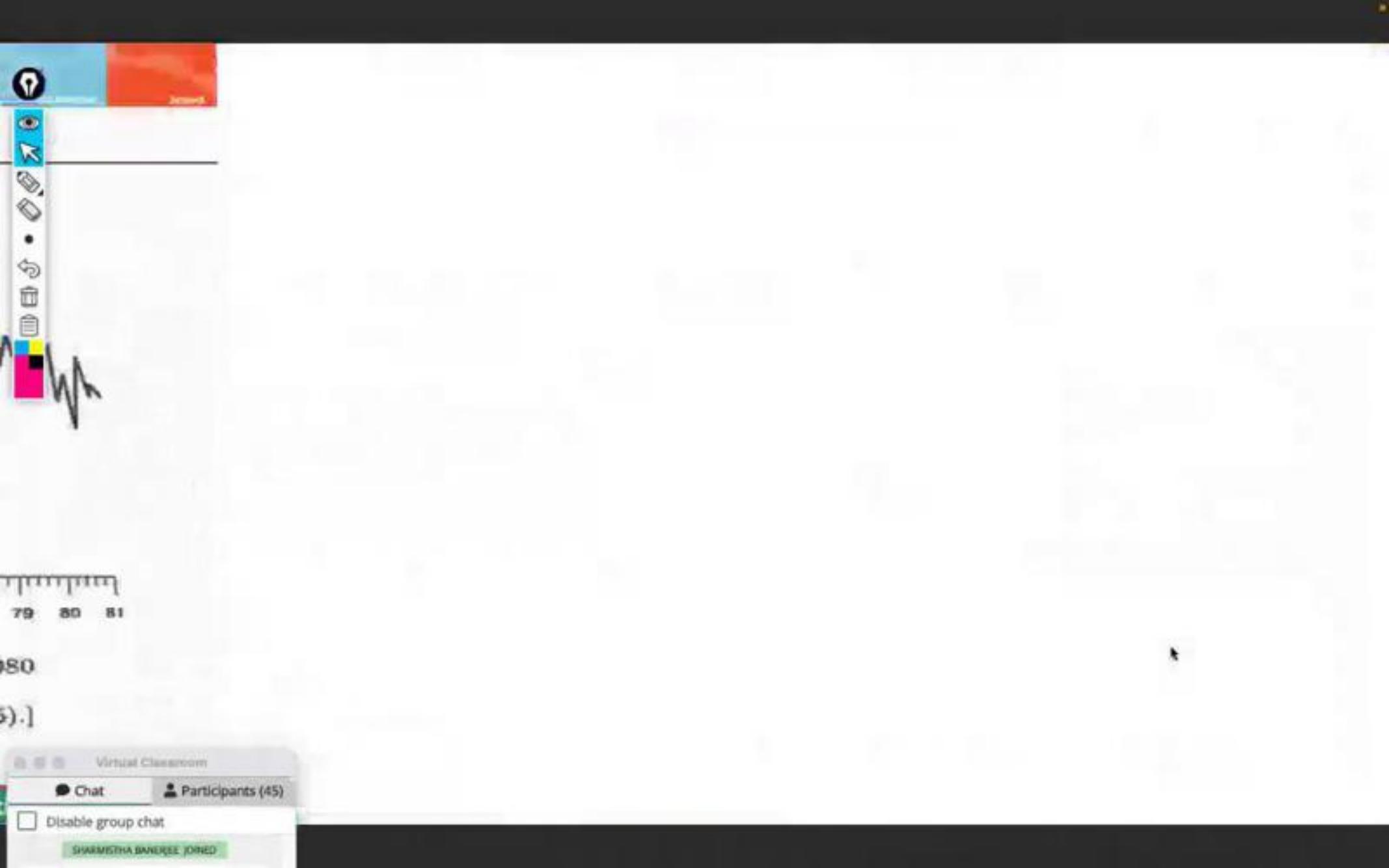
NAGA SUMAN RACHUMALU JOINED

## Time Series → Definition



**Fig 14.3** Short-term movement to and from Australia.

[Source: Immigration update, December, Quarter 1991, Bureau of Immigration Research, Statistics Section, Australia.]



79 80 81

980

5.)]



## Time Series → Decomposition Models

This is an important aspect of time series analysis. It seeks to decompose observed time series into several components. Each of these has a specific type of behaviour. For example, seasonality decomposes

## Time Series → Decomposition Models

This is an important technique for all types of time series analysis, especially for seasonal adjustment. It seeks to construct, from an observed time series, a number of component series (that could be used to reconstruct the original by additions or multiplications) where each of these has a certain characteristic or type of behaviour. For example, time series are usually decomposed into:

## Time Series → Decomposition Models

- $S_t$ : the seasonal component at time  $t$ , reflecting seasonality (seasonal variation). A seasonal pattern exists when a time series is influenced by seasonal factors. Seasonality occurs over a fixed and known period (e.g., the quarter of the year, the month, or day of the week)

- $S_t$ : the seasonal component at time  $t$ , reflecting seasonality (seasonal variation). A seasonal pattern exists when a time series is influenced by seasonal factors. Seasonality occurs over a fixed and known period (e.g., the quarter of the year, the month, or day of the week)

## Time Series → Decomposition Models

$I_t$ : the irregular component (or "noise") at time  $t$ , which describes random, irregular influences. It represents the residuals or remainder of the time series after the other components have been removed.

Virtual Classroom

Chat Participants (45)

Disable group chat

Can we expect problems fitting one & three way ANOVA in exam or just

abraham PhD (Stats)

Former Professor of Statistics | KIMS, B'lore

## Time Series → Decomposition Models

Hence a time series using an additive model can be thought of as:

$$y_t = T_t + C_t + S_t + I_t$$

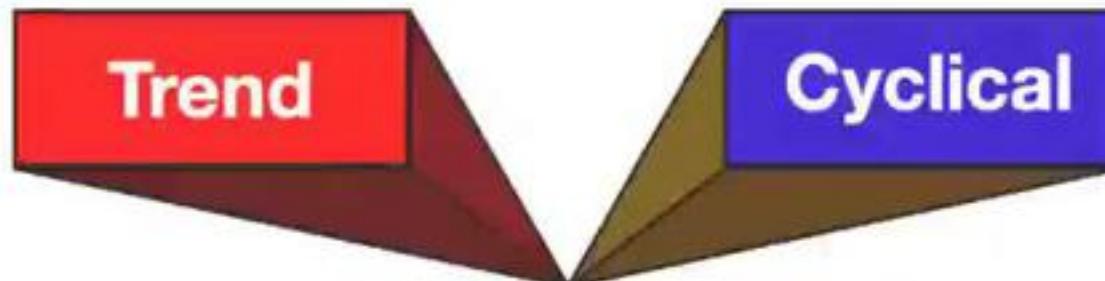
Whereas the multiplicative model can be thought of as:

- An additive model would be used when the variations around the trend do not vary with the level of the time series whereas a multiplicative model would be appropriate if the trend is proportional to the level of the time series.

Sometimes the trend and cyclical components are grouped into one, called the trend-cycle component. The trend-cycle component can just be referred to as the "trend" component, even though it may contain cyclical behaviour. For example, a seasonal decomposition of time series plot decomposes a time series into seasonal, trend and irregular components using loess and plots the components separately, whereby the cyclical component (if present in the data) is included in the "trend" component plot.

## Time Series

### Time Series Components



## Time Series

### Time Series Components

#### Trend:

Response



Mo., Qtr., Yr.

## Time Series

### Time Series Components

#### Trend:

- Persistent, overall upward or downward pattern

Response



Mo., Qtr., Yr.

## Cyclical:

Repeating up & down movements above trend line



#### Cyclical:

Repeating up & down movements above trend line

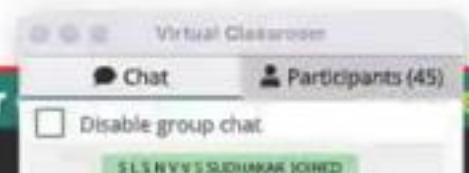
- Due to interactions of factors influencing economy





## Time Series → Smoothing Techniques – Moving Averages

- Appropriate for a time series with a horizontal pattern ie., the data that are stationary.
  - Moving Average (the average of the most recent  $k$  data values forms the forecast for the next period)



Session 13 - 14 Time Series and Forecasting

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years.

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

The graph displays two series: 'Production' (blue line) and '3 - year moving total' (red line). The x-axis represents the years from 1995 to 2004. The y-axis represents production volume in '000 tones, ranging from 20 to 30. The blue line shows the actual production values, which fluctuate between 21 and 27. The red line represents the 3-year moving average, which smooths the data by averaging every three consecutive years. For example, in 1997, the moving average is 22.00, and in 2004, it is 26.33.

Statistical Methods for Data Science  
Time Series → Smoothing Techniques – Moving Averages

Former Professor of Statistics | KJMS, Bangalore

78

Slide 77 of 148 | Created by Prof. Gangaboraiyah B Andanaiah PhD | Last updated on 5 March 2024

Notes Comments Virtual Classroom

(46)

Icons for various applications: Calendar (5), Microsoft Word, Microsoft Excel, Microsoft Powerpoint, Microsoft Access, Microsoft OneDrive, Microsoft Edge, Microsoft Outlook, Microsoft Teams, Microsoft Project, Microsoft Visio, Microsoft Publisher, Microsoft Word, Microsoft Excel, Microsoft Powerpoint, Microsoft Access, Microsoft OneDrive, Microsoft Edge, Microsoft Outlook, Microsoft Teams, Microsoft Project, Microsoft Visio, Microsoft Publisher.

Home Insert Draw Design Transitions Animations Slide Show Review View Table Design Layout

Top Row First Column  
Bottom Row Last Column  
Selected Rows Banded Columns

Shading Borders Effects Quick Styles Pen Color Draw Table Eraser

Statistical Methods for Data Science  
Time Series → Modeling Time Series  
**Modeling Time Series**  
Methods for forecasting a time series fall into two general classes: *smoothing methods* and *regression-based modeling methods*.  
Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components as a basis for building models.

Statistical Methods for Data Science  
Time Series → Smoothing Techniques - Moving Averages  

- Appropriate for a time series with a horizontal pattern i.e., the data are stationary.
- Moving Average (the average of the most recent k data values forms the forecast for the next period)
 
$$\hat{Y}_{t+1} = \frac{\text{sum of most recent data values}}{k}$$

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$$

Statistical Methods for Data Science  
Time Series → Smoothing Techniques - Moving Averages  
 Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years  

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

Statistical Methods for Data Science  
Time Series → Smoothing Techniques - Moving Averages  
 Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years  

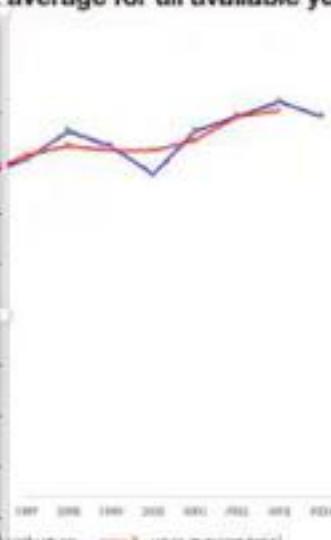
Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	24	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33



Home Insert Draw Design Transitions Animations Slide Show Review View Table Design Layout

Enter Row First Column  
Delete Row Last Column  
Delete Rows Banded Columns

Statistical Methods for Data Science Time Series → Modeling Time Series

**Modeling Time Series**

Methods for forecasting a time series fall into two general classes: smoothing methods and regression-based modeling methods.

Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components as a basis for building models.

Statistical Methods for Data Science Time Series → Smoothing Techniques - Moving Averages

- Appropriate for a time series with a horizontal pattern i.e., the data are stationary.
- Moving Average (the average of the most recent k data values forms the forecast for the next period)
 
$$\bar{Y}_{t+k} = \frac{\text{most recent data values}}{k}$$

$$\bar{Y}_{t+k} = \frac{Y_t + Y_{t+1} + \dots + Y_{t+k-1}}{k}$$

Statistical Methods for Data Science Time Series → Smoothing Techniques - Moving Averages

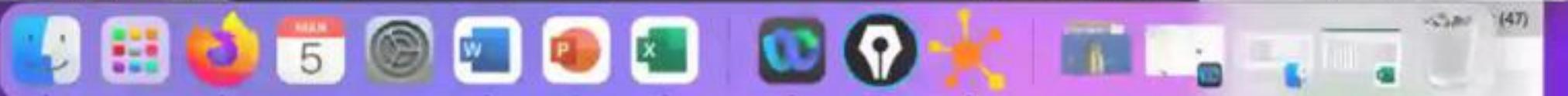
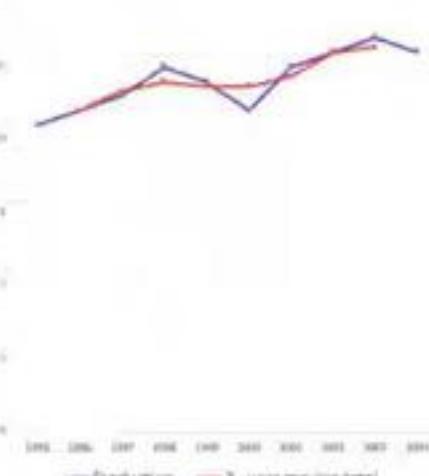
Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33



Session 13 - 14 Time Series and Forecasting

Home Insert Draw Design Transitions Animations Slide Show Review View Table Design Layout

Header Row First Column  
Footer Row Last Column  
Selected Rows Banded Columns

Shading Borders Effects Quick Styles 1 pt Pen Colour Draw Table Eraser

Statistical Methods for Data Science  
Time Series → Modeling Time Series  
**Modeling Time Series**  
Methods for forecasting a time series fall into two general classes: *smoothing methods* and *regression-based modeling methods*.

Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components as a basis for building models.

Statistical Methods for Data Science  
Time Series → Smoothing Techniques – Moving Averages

- Appropriate for a time series with a horizontal pattern i.e., the data are stationary.
- Moving Average (the average of the most recent k data values forms the forecast for the next period)
 
$$F_{t+1} = \frac{\text{most recent } k \text{ data values}}{k}$$

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$$

Statistical Methods for Data Science  
Time Series → Smoothing Techniques – Moving Averages

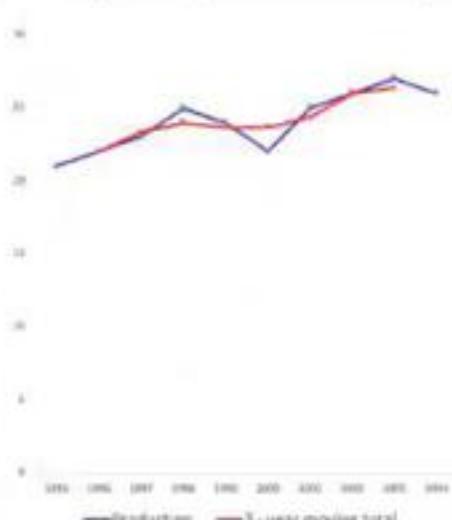
Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33



**Statistical Methods for Data Science**

**Time Series** → Modeling Time Series

**Modeling Time Series**

Methods for forecasting a time series fall into two general classes: *smoothing methods* and *regression-based modeling methods*.

- Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components as a basis for building models.

**Statistical Methods for Data Science**

**Time Series** → Smoothing Techniques – Moving Averages

- Appropriate for a time series with a horizontal pattern i.e., the data are stationary.
- Moving Average (the average of the most recent k data values forms the forecast for the next period)
 
$$F_{t+1} = \frac{\text{most recent data values}}{k}$$

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$$

**Statistical Methods for Data Science**

**Time Series** → Smoothing Techniques – Moving Averages

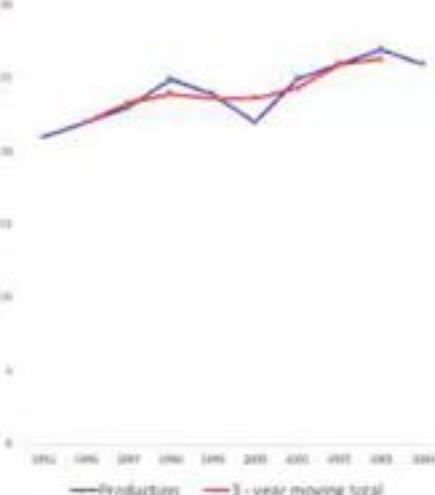
Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33



Slide 77

Prof.Gangaboraiyah PhD (IITM)  
Visiting Classroom

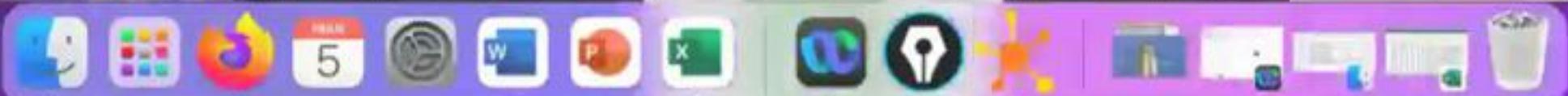
Former Professor of Statistics | KIMS, Bangalore

Chat Participants (47)

Disable group chat

Notes Comments

28%



Session 13 - 14 Time Series and Forecasting

**Statistical Methods for Data Science**

**Time Series** → **Modeling Time Series**

**Modeling Time Series**

- Methods for forecasting a time series fall into two general classes: **smoothing methods** and **regression-based modeling methods**.
- Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components, as a basis for building models.

**Statistical Methods for Data Science**

**Time Series** → **Smoothing Techniques – Moving Averages**

- Appropriate for a time series with a horizontal pattern i.e., the data are stationary.
- Moving Average (the average of the most recent k data values forms the forecast for the next period)
 
$$F_{t+1} = \frac{\text{most recent data values}}{k}$$

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$$

**Statistical Methods for Data Science**

**Time Series** → **Smoothing Techniques – Moving Averages**

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

**Statistical Methods for Data Science**

**Time Series** → **Smoothing Techniques – Moving Averages**

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

The graph displays the production volume over time. The x-axis represents the years from 1995 to 2004. The y-axis represents the production volume in thousands of tones, ranging from 20 to 30. The red line represents the actual production data, which shows some fluctuations. The blue line represents the 3-year moving average, which smooths out the fluctuations and shows a general upward trend.

**Former Professor of Statistics | KIMS, B'lore**



# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

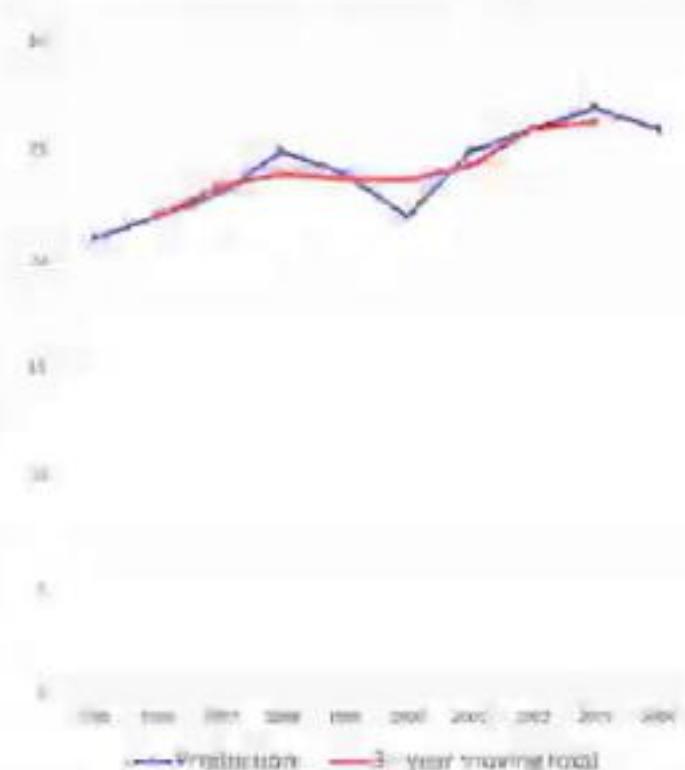
Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33

# Statistical Methods for Data Science

## Series → Smoothing Techniques – Moving Averages

The data shows production volume (in '000 tones). Compute 3-year moving average for all available years

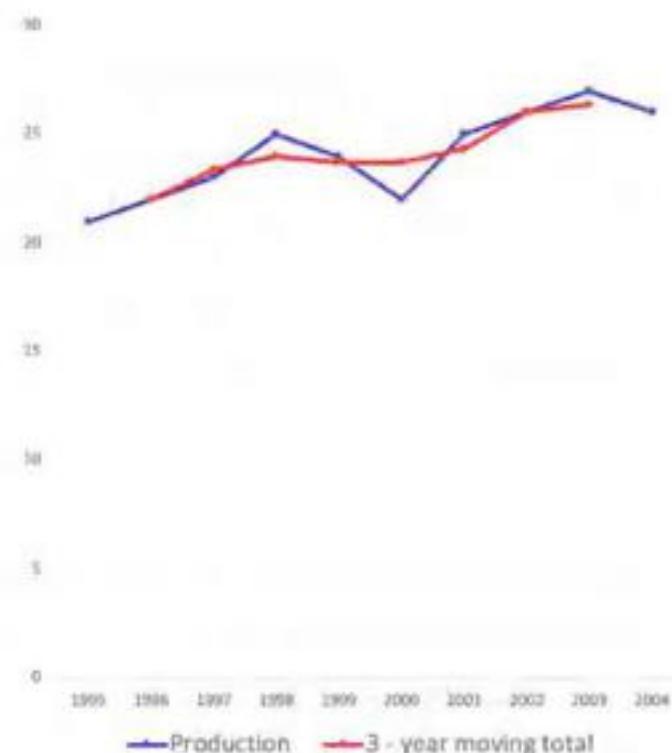
Year	Production	3 - year moving total	3 - year moving average
1996	21		
1997	22		
1998	23	66	22.00
1999	25	70	23.33
2000	24	72	24.00
2001	22	71	23.67
2002	25	71	23.67
2003	26	73	24.33
2004	27	78	26.00
2005	26	79	26.33



## Time Series → Smoothing Techniques – Moving Averages

The following data shows production volume (in '000 tones). Compute 3-year moving average for all available years

Year	Production	3 - year moving total	3 - year moving average
1995	21		
1996	22		
1997	23	66	22.00
1998	25	70	23.33
1999	24	72	24.00
2000	22	71	23.67
2001	25	71	23.67
2002	26	73	24.33
2003	27	78	26.00
2004	26	79	26.33



- Exponential (Weighted average of all the past time series values with exponentially decreasing importance in the forecast)

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

- $\alpha=1 \Rightarrow$  Forecast value is just the previous actual value. A naïve forecast.
- $\alpha=0 \Rightarrow$  Current Forecast value is just the previous forecast value.
- How to choose value of  $k$  and  $\alpha$ ?

# Statistical Methods for Data Science

Introduction

Applications

Index

## Time Series → Smoothing Techniques – Exponential

- Exponential (Weighted average of all the past time series values with exponentially decreasing importance in the forecast)

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

- $\alpha=1 \Rightarrow$  Forecast value is just the previous actual value. A naïve forecast.
- $\alpha=0 \Rightarrow$  Current Forecast value is just the previous forecast value.
- How to choose value of k and  $\alpha$ ?





e. A

cast

KIMS, Blore

Virtual Classroom

Chat Participants (44)

Disable group chat

WASIM AKRAM JOINED

# Statistical Methods for Data Science

INTERACTIVE

INTERACTIVE

READ

## Time Series

### Smoothing Techniques – Exponential

Virtual Classroom

- Chat
- Participants (44)
- Disable group chat

Chinnam Gopal Chari...  
82.26

Gautam Raj  
82.26

Alok  
82.26

Balajee R.  
sir.. we are using the that year actual  
to come up with predictor ?

Gautam Raj  
@Balajee Yt is from previous year

Alok  
should we not take 79

79\*0.2

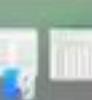
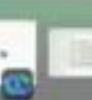
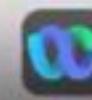
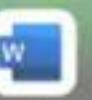
ZENATH BIR JOINED

SLSN VVS Sudhar  
I think we need to take  
79\*02+8\*82.26

Writing to everyone  
Write Message

the	Quarter/ Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
3	1/2003	82.50			
	2/2003	81.30	82.5	82.5	82.5
	3/2003	81.30	82.26	82.02	81.66
	4/2003	79.00	81.61	80.81	79.80
	1/2004	76.60	80.61	79.13	77.56
	2/2004	78.00	80.09	78.68	77.87
	3/2004	78.40	79.75	78.57	78.24
	4/2004	78.00	79.40	78.34	78.07
	1/2005	78.80	79.28	78.52	78.58
	2/2005	78.70	79.16	78.59	78.66
	3/2005	78.40	79.01	78.52	78.48
	4/2005	80.00	79.21	79.11	79.54
	1/2006	80.70	79.51	79.75	80.35
	2/2006	80.70	79.75	80.13	80.60
	3/2006	80.80	79.96	80.40	80.74
	MSE	2.70	1.04	0.26	

$$- \alpha) F_t$$



Time-series data

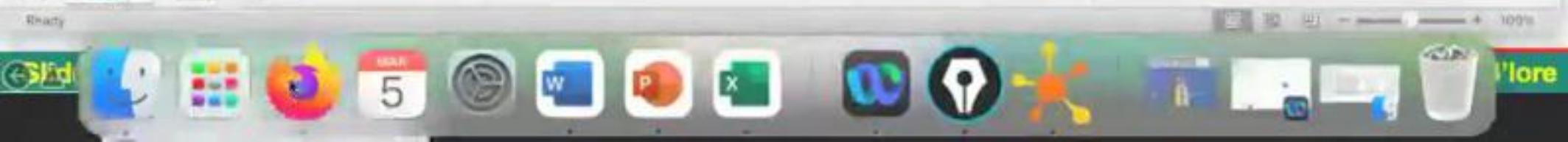
an recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

1593

- swing data gives the cargo exports  $y_t$  (in units of  $10^3$  metric tons) at the principal port in India for a period of 20 years
- covariance, autocorrelation and auto regression

	lag 1 ( $y_1$ )	lag 2 ( $y_2 = \text{lag3 } (y_1 = \text{lag 1 } (y_4 = \text{lag 5 } (y_1 = yt-m))$ )	lag 1-m	$y_{t-m}$	$y_{2-m}$	$y_{3-m}$	$y_{4-m}$	$y_{5-m}$	$(yt-m)(y_{t-m})$	$(yt-m)(y_{t-m})(y(yt-m))$	$(yt-m)(y(yt-m)(y(yt-m)(y5-m)))$
1230	1345	1382	1416	1593	1802	-1010.2	-895	-858	-824	-647	-438.2
1345	1382	1416	1593	1802	1817	-895.2	-858	-824	-647	-438	-423.2
1382	1416	1593	1802	1817	1995	-858.2	-824	-647	-438	-423	-245.2
1416	1593	1802	1817	1995	2212	-824.2	-647	-438	-423	-245	-28.2
1593	1802	1817	1995	2212	2607	-647.2	-438	-423	-245	-28	366.8
1802	1817	1995	2212	2607	2698	-438.2	-423	-245	-28	367	457.8
1817	1995	2212	2607	2698	2703	-423.2	-245	-28	367	458	462.8
1995	2212	2607	2698	2703	2738	-245.2	-28	367	458	463	497.8
2212	2607	2698	2703	2738	2810	-28.2	367	458	463	498	569.8
2607	2698	2703	2738	2810	2632	366.8	458	463	498	570	391.8
2698	2703	2738	2810	2632	2548	457.8	463	498	570	392	307.8
2703	2738	2810	2632	2548	2881	462.8	498	570	392	308	640.8
2738	2810	2632	2548	2881	2725	497.8	570	392	308	641	484.8
2810	2632	2548	2881	2725	2911	569.8	392	308	641	485	670.8
2632	2548	2881	2725	2911	2759	391.8	308	641	485	671	518.8



Time series data

Recent changes were saved. Do you want to continue working where you left off?

**Format Cells**

Number Alignment Font Border Fill Protection

Horizontal Alignment: Center Vertical Alignment: Bottom Orientation: Text Text Degrees: 0

Text control: Wrap text Merge cells

Cancel OK

Quarter/year			
Jan-03	-0.2)	F <sub>1</sub> ( $\alpha=0.4$ )	F <sub>1</sub> ( $\alpha=0.7$ )
Feb-03			
Mar-03			
Apr-03			
Jan-04			
Feb-04			
Mar-04			
Apr-04			
Jan-05			
Feb-05			
Mar-05			
Apr-05			
Jan-06			
Feb-06			
Mar-06			

Time series data

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format

Calibri (Body) 22 A A+ = Wrap Text General Conditional Formatting Cell Styles Insert Delete Format

Auto-sum Fill Clear Sort & Filter Find & Select

Jan-03	82.5	NA	NA	NA		1	82.5	NA	NA	NA														
Feb-03	81.3	82.5	82.5	82.5		2	81.3	82.5	82.5	82.5	82.5	82.5												
Mar-03	81.3	82.26	82.02	81.66		3	81.3	82.26	82.02	81.66														
Apr-03	79	81.61	80.81	79.8		4	79	82.068	81.732	81.408														
Jan-04	76.6	80.61	79.13	77.56		5	76.6	81.4544	80.6392	79.7224														
Feb-04	78	80.09	78.68	77.87		6	78	80.48352	79.02352	77.53672														
Mar-04	78.4	79.75	78.57	78.24		7	78.4	79.98682	78.61411	77.86102														
Apr-04	78	79.4	78.34	78.07		8	78	79.66945	78.52847	78.2383														
Jan-05	78.8	79.28	78.52	78.58		9	78.8	79.33556	78.31708	78.07149														
Feb-05	78.7	79.16	78.59	78.66		10	78.7	79.22845	78.51025	78.58145														
Mar-05	78.4	79.01	78.52	78.48		11	78.4	79.12276	78.58615	78.66443														
Apr-05	80	79.21	79.11	79.54		12	80	78.97821	78.51169	78.47933														
Jan-06	80.7	79.51	79.75	80.35		13	80.7	79.18257	79.10701	79.5438														
Feb-06	80.7	79.75	80.13	80.6		14	80.7	79.48605	79.74421	80.35314														
Mar-06	80.8	79.96	80.4	80.74		15	80.8	79.72884	80.12652	80.59594														
3/2006						16		79.94307	80.39591	80.73878														



# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

Done the analysis on the previous problem for 3 different values of  $\alpha$

For  $\alpha=0.7$ , it is better

The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed

Formula can be rearranged as

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

Quarter/ Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.5	82.5	82.5
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
	MSE	2.70	1.04	0.26

Virtual Classroom

Chat

Participants (49)

Disable group chat

(Stats)

Former Professor of Statistics | KIMS, B'lore

## Statistical Methods for Data Science

Time Series → Modeling Time Series

## Modeling Time Series

Methods for forecasting a time series fall into two general classes: smoothing methods and regression-based modeling methods.

Although the smoothing methods do not explicitly use the time series components, it is a good idea to keep them in mind. The regression models explicitly estimate the components as a basis for building models.

## Statistical Methods for Data Science

Time Series → Smoothing Techniques – Moving Averages

- appropriate for a time series with a nonstationary pattern, the data that are stationary
- Moving Average does not care about the trend term. Data values forms the forecast for the next point

$$\text{Forecast} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$$

$$F_t = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$$

## Statistical Methods for Data Science

Time Series → Smoothing Techniques – Exponential Smoothing

Exponential Smoothing is a time series forecasting technique that uses a weighted average of past observations to predict future values.

The formula for exponential smoothing is:

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

where  $F_t$  is the forecast for time  $t$ ,  $Y_t$  is the actual value at time  $t$ , and  $\alpha$  is the smoothing factor.The smoothing factor  $\alpha$  determines the weight given to the most recent observation.A higher value of  $\alpha$  places more weight on the most recent observation, while a lower value places more weight on the previous forecasts.The initial forecast  $F_1$  is typically set to the first observed value  $Y_1$ .

The exponential smoothing formula can be rearranged as:

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

The forecast for time  $t+1$  is calculated by multiplying the current observation  $Y_t$  by the smoothing factor  $\alpha$  and adding the previous forecast  $F_t$  multiplied by  $(1-\alpha)$ .The forecast for time  $t+1$  is then used as the forecast for time  $t+2$ , and so on.

The exponential smoothing formula is:

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

The forecast for time  $t+1$  is calculated by multiplying the current observation  $Y_t$  by the smoothing factor  $\alpha$  and adding the previous forecast  $F_t$  multiplied by  $(1-\alpha)$ .The forecast for time  $t+1$  is then used as the forecast for time  $t+2$ , and so on.

The exponential smoothing formula is:

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

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed
- Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.5	82.5	82.5
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006				

79 Statistical Methods for Data Science Time Series → Smoothing Techniques – Exponential

- Exponential (Weighted average of all the past time series values with exponentially decreasing importance in the forecast)
 
$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$
- $\alpha=1$  Forecast value is just the previous actual value. A naive forecast.
- $\alpha=0$  Current Forecast value is just the previous forecast value.
- How to choose value of  $k$  and  $\alpha$ ?

80 Statistical Methods for Data Science Time Series → Smoothing Techniques – Exponential

- Carry the analysis on the previous problem for 5 different values of  $\alpha$
- For  $\alpha=1$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed
- Formula can be rearranged as
 
$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

81 Statistical Methods for Data Science Time Series → Smoothing Techniques – Exponential

- Carry the analysis on the previous problem for 5 different values of  $\alpha$
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed
- Formula can be rearranged as
 
$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

## Statistical Methods for Data Science

### Time Series → Smoothing Techniques – Exponential

Quarter/Year	Utilization (%)	$F_1(\alpha=0.2)$	$F_1(\alpha=0.4)$	$F_1(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	82.07	81.73	81.41
1/2004	76.60	81.45	80.64	79.72
2/2004	78.00	80.48	79.02	77.54
3/2004	78.40	79.99	78.61	77.86
4/2004	78.00	79.67	78.53	78.24
1/2005	78.80	79.34	78.32	78.07
2/2005	78.70	79.23	78.51	78.58
3/2005	78.40	79.12	78.59	78.66
4/2005	80.00	78.98	78.51	78.48
1/2006	80.70	79.18	79.11	79.54
2/2006	80.70	79.49	79.74	80.35
3/2006	80.80	79.73	80.13	80.60
4/2006				



Statistical Methods for Data Science  
Time Series → Smoothing Techniques – Exponential

- Exponentially Weighted average of all the past time series values with exponentially decreasing importance in the forecast.

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

- $\alpha=1$ : Forecast value is just the previous actual value. A naive forecast.
- $\alpha=0$ : Current Forecast value is just the previous forecast value.
- How to choose value of  $k$  and  $\alpha$ ?

Statistical Methods for Data Science  
Time Series → Smoothing Techniques – Exponential

Date	Actual	Smoothed	Actual	Smoothed
1/2003	82.50	82.50	81.30	81.30
2/2003	81.30	82.50	81.30	82.50
3/2003	81.30	82.26	81.30	82.02
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74

Statistical Methods for Data Science  
Time Series → Smoothing Techniques – Exponential

Date	Actual	Smoothed	Actual	Smoothed
1/2003	82.50	82.50	80.50	80.50
2/2003	81.30	82.27	81.30	82.27
3/2003	81.30	82.04	81.30	82.04
4/2003	79.00	81.44	79.70	79.70
1/2004	76.60	80.80	76.22	77.22
2/2004	78.00	80.00	78.01	79.01
3/2004	78.40	79.77	79.44	79.44
4/2004	78.00	79.27	78.22	78.22
1/2005	78.80	79.50	78.67	79.67
2/2005	78.70	79.25	78.77	79.77
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74

Statistical Methods for Data Science  
Time Series

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$ .
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed.
- Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.5	82.5	82.5
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74



Session 13 - 14 Time Series and Forecasting

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

Done the analysis on the previous problem for 3 different values of  $\alpha$ . For  $\alpha=0.7$ , it is better. The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.

For  $\alpha=0.7$ , it is better. The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.

The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.

Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t (\alpha=0.2)$	$F_t (\alpha=0.4)$	$F_t (\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
3/2006	-	79.94	80.40	80.74

80 | Prof. Gangaboralih B Andanalath PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboraiyah B Andanalal PhD Sat 5 Mar 7:03 PM

Home Insert Draw Page Layout Formulas Data Review View

Time series data

Cut Copy Paste Format

Calibri (Body) 22 A+ A- Wrap Text Number

Merge & Center Conditional Formatting Insert Delete Format

Sort & Filter Find & Select Yes No

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

f(x) = [0.2 \* H16] + [0.8 \* I16]

	Jan-03	82.5	1	82.5	NA	NA
	Feb-03	81.3	2	81.3	82.50	82.50
	Mar-03	81.3	3	81.3	82.26	82.02
	Apr-03	79	4	79	82.07	81.73
	Jan-04	76.6	5	76.6	81.45	80.64
	Feb-04	78	6	78	80.48	79.02
	Mar-04	78.4	7	78.4	79.99	78.61
	Apr-04	78	8	78	79.67	78.53
	Jan-05	78.8	9	78.8	79.34	78.32
	Feb-05	78.7	10	78.7	79.23	78.51
	Mar-05	78.4	11	78.4	79.12	78.59
	Apr-05	80	12	80	78.98	78.51
	Jan-06	80.7	13	80.7	79.18	79.11
	Feb-06	80.7	14	80.7	79.49	79.74
	Mar-06	80.8	15	80.8	79.73	80.13
	3/2006		16	79.94	80.40	80.74

Chat Participants (50)

- Disable group chat
- yt is the actual value?
- Sai Kumar K.C
- Sir, Y value is % Utilization. Right?
- Balajeet R
- without F1...what will we do for F2
- ahh...oke
- SRIVATSAN ER JOINED
- got it
- thanks
- PRASHANT SINGH JOINED
- MALAVIKA SHAMESH JOINED
- Sunit Kumar Pradhan
- Sir - could you please share the column H values here
- BHATT NAXIN SURESHKUMAR JOINED
- B C HEMANTH KUMAR JOINED

Showing 10 everyone Write Message

Microsoft Excel



Session 13 – 14 Time Series and Forecasting

# Statistical Methods for Data Science

## Time Series ➡ Smoothing Techniques – Exponential

Done the analysis on the previous problem for 3 different values of  $\alpha$ .  
For  $\alpha=0.7$ , it is better

The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed.

Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t (\alpha=0.2)$	$F_t (\alpha=0.4)$	$F_t (\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.86
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74

80 Prof. Gangaboraiyah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

Statistical Methods for Data Science  
Time Series ➡ Smoothing Techniques – Exponential

- Exponential Weighted Average of all the past time series values with exponential weight
- Alpha Forecast value X<sub>t+1</sub> new forecast
- Alpha Current Forecast value
- New to choose value of alpha

Virtual Classroom

Chat Participants (50)

Disable group chat

Sai Kumar K C  
Sir, Y value is % Utilization. Right?

Galaxy R  
without F1...what will we do for F2

ahh...oke

SRIVATSAN V R JOINED

got it

thanks

PRASHANT SINGH JOINED

MALAVIKA SHAMESH JOINED

Sunit Kumar Pradhan  
Sir - could you please share the column H values here

BHATT NAYAN SURISH KUMAR JOINED

S C HEMANTH KUMAR JOINED

No sir please ignore I copied it

Writing to everyone  
Write Message

Slide 28 of 346 English



Section 12-14 Time Series and Forecasting

© 2010 Pearson Education, Inc.

100



Insert  
Cut  
Copy \*  
Format

Layout  Helvetica

Statistical Methods for Data Science

### Time Series with Smoothing Techniques

- Exponential (Weighted average of all the past time series values with exponentially decreasing importance in the forecast)
 
$$F_{t+1} = \alpha Y_t + (1-\alpha) F_t$$
  - $\alpha=1$  Forecast value is just the previous actual value. A naive forecast.
  - $\alpha=0$  Current Forecast value is just the previous forecast value.
  - How to choose value of  $\alpha$  &  $k$ ?

Statistical Methods for Data Science

Time Series and Smoothing Techniques - Exponential

- |    | Time      | Day | Y = Sales | $\hat{Y}_{t+1}$ = Forecast | $\epsilon_t$ = Residual |
|----|-----------|-----|-----------|----------------------------|-------------------------|
| 1  | 1/1/2010  | 1   | 20.00     | 20.00                      | 0.00                    |
| 2  | 1/2/2010  | 2   | 21.00     | 20.00                      | -1.00                   |
| 3  | 1/3/2010  | 3   | 20.00     | 21.00                      | 1.00                    |
| 4  | 1/4/2010  | 4   | 21.00     | 20.00                      | -1.00                   |
| 5  | 1/5/2010  | 5   | 20.00     | 21.00                      | -1.00                   |
| 6  | 1/6/2010  | 6   | 21.00     | 20.00                      | -1.00                   |
| 7  | 1/7/2010  | 7   | 20.00     | 21.00                      | -1.00                   |
| 8  | 1/8/2010  | 8   | 21.00     | 20.00                      | -1.00                   |
| 9  | 1/9/2010  | 9   | 20.00     | 21.00                      | -1.00                   |
| 10 | 1/10/2010 | 10  | 21.00     | 20.00                      | -1.00                   |
| 11 | 1/11/2010 | 11  | 20.00     | 21.00                      | -1.00                   |
| 12 | 1/12/2010 | 12  | 21.00     | 20.00                      | -1.00                   |
| 13 | 1/13/2010 | 13  | 20.00     | 21.00                      | -1.00                   |
| 14 | 1/14/2010 | 14  | 21.00     | 20.00                      | -1.00                   |
| 15 | 1/15/2010 | 15  | 20.00     | 21.00                      | -1.00                   |
| 16 | 1/16/2010 | 16  | 21.00     | 20.00                      | -1.00                   |
| 17 | 1/17/2010 | 17  | 20.00     | 21.00                      | -1.00                   |
| 18 | 1/18/2010 | 18  | 21.00     | 20.00                      | -1.00                   |
| 19 | 1/19/2010 | 19  | 20.00     | 21.00                      | -1.00                   |
| 20 | 1/20/2010 | 20  | 21.00     | 20.00                      | -1.00                   |
| 21 | 1/21/2010 | 21  | 20.00     | 21.00                      | -1.00                   |
| 22 | 1/22/2010 | 22  | 21.00     | 20.00                      | -1.00                   |
| 23 | 1/23/2010 | 23  | 20.00     | 21.00                      | -1.00                   |
| 24 | 1/24/2010 | 24  | 21.00     | 20.00                      | -1.00                   |
| 25 | 1/25/2010 | 25  | 20.00     | 21.00                      | -1.00                   |
| 26 | 1/26/2010 | 26  | 21.00     | 20.00                      | -1.00                   |
| 27 | 1/27/2010 | 27  | 20.00     | 21.00                      | -1.00                   |
| 28 | 1/28/2010 | 28  | 21.00     | 20.00                      | -1.00                   |
| 29 | 1/29/2010 | 29  | 20.00     | 21.00                      | -1.00                   |
| 30 | 1/30/2010 | 30  | 21.00     | 20.00                      | -1.00                   |
| 31 | 1/31/2010 | 31  | 20.00     | 21.00                      | -1.00                   |

Statistical Methods for Data Science

Time Series w/ Smoothing Techniques - Exponential

- |                          | Mean  | SD    | SE   | $t_{\text{crit}}$ |
|--------------------------|-------|-------|------|-------------------|
| Control                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group A                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group B                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group C                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group D                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group E                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group F                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group G                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group H                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group I                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group J                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group K                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group L                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group M                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group N                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group O                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group P                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group Q                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group R                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group S                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group T                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group U                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group V                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group W                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group X                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group Y                  | 21.31 | 20.00 | 4.62 | 2.77              |
| Group Z                  | 21.31 | 20.00 | 4.62 | 2.77              |
| $F_{\text{crit}} = 9.27$ | 21.31 | 20.00 | 4.62 | 2.77              |

Statistical Methods for Data Science Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$
  - For  $\alpha=0.7$ , it is better
  - The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed
  - Formula can be rearranged as

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

Quarter/ Year	Utilization (%)	$F_1(a=0.2)$	$F_1(a=0.4)$	$F_1(a=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
3/2006	-	79.94	80.40	80.74



Session 13 - 14 Time Series and Forecasting

Search in Presentation

Home Insert Draw Design Transitions Animations Slide Show Review View

Cut Copy New Slide Layout Reset Section

Font Helvetica Neue Size 22

Text Box Picture Shapes Quick Styles Shape Outline

Statistical Methods for Data Science	
Time Series	Smoothing Techniques - Exponential
<ul style="list-style-type: none"> <li>Exponential (Weighted average of all the past time series values with exponentially decreasing importance in the forecast) <math>F_{t+1} = \alpha Y_t + (1-\alpha)F_t</math></li> <li><math>\alpha=1</math> Forecast value is just the previous actual value. A naive forecast.</li> <li><math>\alpha=0</math> Current Forecast value is just the previous forecast value.</li> <li>How to choose value of k and <math>\alpha</math>?</li> </ul>	
Done the analysis for 5 different values of $\alpha$ .	
For $\alpha=0.7$ , it is better.	
The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.	
Formula can be rearranged as	
$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$	Year 2003-2006

Statistical Methods for Data Science	
Time Series	Smoothing Techniques - Exponential
<ul style="list-style-type: none"> <li>Done the analysis for 5 different values of <math>\alpha</math>.</li> <li>For <math>\alpha=0.7</math>, it is better.</li> <li>The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.</li> <li>Formula can be rearranged as</li> </ul>	
$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$	Year 2003-2006

Statistical Methods for Data Science	
Time Series	Smoothing Techniques - Exponential
<ul style="list-style-type: none"> <li>Done the analysis for 5 different values of <math>\alpha</math>.</li> <li>For <math>\alpha=0.7</math>, it is better.</li> <li>The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.</li> <li>Formula can be rearranged as</li> </ul>	
$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$	Year 2003-2006

Statistical Methods for Data Science			
Time	Actual	Smoothed	Error
1/2003	82.50	82.50	0.00
2/2003	81.30	82.50	-1.20
3/2003	81.30	82.26	-0.96
4/2003	79.00	81.61	-2.61
1/2004	76.60	80.61	-3.99
2/2004	78.00	80.09	-2.09
3/2004	78.40	79.75	-0.75
4/2004	78.00	79.40	-1.40
1/2005	78.80	79.28	-0.52
2/2005	78.70	79.16	-0.46
3/2005	78.40	79.01	-0.61
4/2005	80.00	79.21	0.79
1/2006	80.70	79.51	1.19
2/2006	80.70	79.75	1.15
3/2006	80.80	79.96	0.84
4/2006	80.40	80.74	-0.34

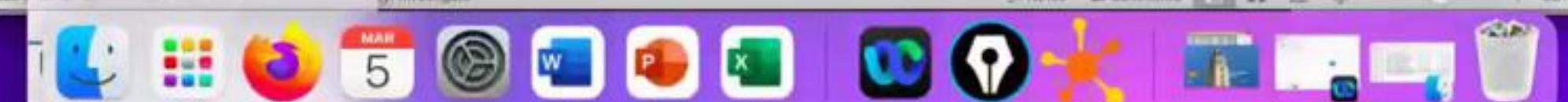
# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$ .
- For  $\alpha=0.7$ , it is better.
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed.
- Formula can be rearranged as

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

Quarter/ Year	Utilization (%)	$F_t (\alpha=0.2)$	$F_t (\alpha=0.4)$	$F_t (\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74



Session 13 - 14 Time Series and Forecasting

Search in Presentation Share

Home Insert Draw Design Transitions Animations Slide Show Review View

Cut Copy Paste Format New Slide Layout Reset Section

Font Helvetica Neue Size 22

Text Box Picture Shapes Quick Styles Shape Outline

82 Statistical Methods for Data Science

Time Series → Smoothing Techniques - Exponential

- Trend-line forecasting (TTF)
- Exponential smoothing (ES)
- The trend line is linear
- The exponential smoothing does not make assumptions about what lies ahead.
- No need to initial values
- One-step-ahead forecast
- Exponential smoothing

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha) \hat{Y}_t$$

83 Statistical Methods for Data Science

Time Series → Smoothing Techniques - Exponential

- Decreases after increasing the forecast error of a prediction by taking into account the previous error.
- Additive Model:  $\hat{Y}_t = F_t + E_t$
- Multiplicative Model:  $\hat{Y}_t = F_t \times E_t$
- Multiplicative models can be converted to additive models by taking log.
- Trend and seasonal components can be decomposed using smoothing techniques.
- Exponential smoothing is suitable with constant variance and no seasonal fluctuations.
- Exponential smoothing is suitable for short-term forecasts.
- Additive method for decomposing time series by aligning nonstationarity e.g. Differencing.

84 Statistical Methods for Data Science

Time Series → Decomposition view

# Statistical Methods for Data Science

## Time Series → Decomposition view



Session 13 - 14 Time Series and Forecasting

Home Insert Draw Design Transitions Animations Slide Show Review View

Cut Copy Format Layout Reset Section

Picture Shapes Text Box Arrange Cut & Paste Copy & Paste

Statistical Methods for Data Science					
Time Series $\Rightarrow$ Smoothing Techniques – Exponential					
- Done the analysis on the previous problem for 3 different values of $\alpha$					
- For $\alpha=0.2$ , it is better					
$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$					

Statistical Methods for Data Science					
Time Series $\Rightarrow$ Smoothing Techniques – Exponential					
- Objective is to estimate the original time series as a combination of long term trend and seasonality					
- Additive Model: $X_t = F_t + S_t + \epsilon_t$					
- Multiplicative Model: $X_t = F_t \times S_t \times \epsilon_t$					
- Multiplicative model can be converted to additive model by taking log. Other models are also possible					
- The trend and seasonality can be incorporated using smoothing and regression methods.					
- Exponential smoothing is suitable with constant variance and no seasonality. Recommended for short-term forecast.					
- Another method for visualizing the time series is by doing transformation, e.g. Differencing					

Statistical Methods for Data Science					
Time Series $\Rightarrow$ Decomposition view					

# Statistical Methods for Data Science

## Time Series $\Rightarrow$ Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial k values are not smoothed
- Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t (\alpha=0.2)$	$F_t (\alpha=0.4)$	$F_t (\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74



**Statistical Methods for Data Science**

**Time Series** → **Smoothing Techniques – Exponential**

- Objective is to estimate the overall time series as a combination of long term trend and seasonality.
- Additive Model:  $X_t = F_t + S_t + \epsilon_t$
- Multiplicative Model:  $X_t = F_t \times S_t \times \epsilon_t$
- Multiplicative model can be converted to additive model by taking log. Other models are also possible.
- The trend and seasonality can be decomposed using smoothing and regression methods.
- Exponential smoothing is suitable with constant variance and no seasonality. Recommended for short-term forecast.
- Another method for stationarizing the time series is by doing transformation .eg. Differencing.

**Statistical Methods for Data Science**

**Time Series** → **Decomposition view**

**Statistical Methods for Data Science**

**Time Series** → **Decomposition view**

- Take the time series and by decomposing it into fitting (regression) and error.
- Residual represent the measurement i.e. unmeasurable data.
- For Additive Adjustment: Look at an periods in a given-type say, Five. Compute when there is a quantity, so as to neglect period when the data is negative & compute an average deviation of the actual values from the expected or fitted values in those periods. The average can then be added to the next to update the measurements.
- For Multiplicative adjustment: Instead of calculating the average deviation, compute an average ratio, also called seasonal indices, of the actual values to the expected or fitted values in fixed periods. The ratios are then used as multiplier to adjust the measurements.
- Periods are selected very conveniently the future will have nothing to do with adjusting the previous components; all the past can be according to your choice.

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Objective is to estimate the overall time series as a combination of long term trend and seasonality
  - Additive Model:  $X_t = F_t + S_t + \epsilon_t$
  - Multiplicative Model:  $X_t = F_t \times S_t \times \epsilon_t$
  - Multiplicative model can be converted to additive model by taking log. Other models are also possible
- The trend and seasonality can be decomposed using smoothing and regression methods
- Exponential smoothing is suitable with constant variance and no seasonality. Recommended for short-term forecast.
- Another method for stationarizing the time series is by doing transformation .eg. Differencing



122 Statistical Methods for Data Science Time Series ↗ Which performance should be used?

- Choice of the modeler
- Advantages
  - MSE gives greater weight to larger deviations (which could result from outliers)
  - MAD gives less weight to larger deviations
  - MAPE gives less overall weight to a large deviation if the time series value is large
  - LAD tells us if all deviations fall below some threshold value
- Although we illustrate all 4 techniques, in general focus will be on MSE and MAD.

123 Statistical Methods for Data Science Time Series ↗ Smoothing Techniques – Moving Averages

The Moving Average basically filters out rapid fluctuations, i.e. high frequency noise. Thus it acts as a low-pass filter.

We can choose the index of the MA filter by trial and error. Comparing the Forecast Accuracy for Index selection.

For our eg., we find that MSE for the 2-Chr filter is better.

124 Statistical Methods for Data Science Time Series ↗ Smoothing Techniques – Moving Averages

Date	Unfilled	Ch1	ASR1	PF1	Ch2	ASR2	PF2	ASR3	PF3	ASR4	PF4
Jan-03	90.5										
Feb-03	92.5	91.8	92.7	92.6	91.5	91.4	91.6	91.6	91.6	91.6	91.6
Mar-03	93.5	92.4	93.2	93.1	92.3	92.2	92.4	92.4	92.4	92.4	92.4
Apr-03	94.5	93.1	93.9	93.8	92.9	92.8	93.0	93.0	93.0	93.0	93.0
May-03	95.5	93.8	94.6	94.5	93.7	93.6	93.8	93.8	93.8	93.8	93.8
Jun-03	96.5	94.5	95.3	95.2	94.4	94.3	94.5	94.5	94.5	94.5	94.5
Jul-03	97.5	95.2	96.0	95.9	95.1	95.0	95.2	95.2	95.2	95.2	95.2
Aug-03	98.5	95.9	96.7	96.6	95.8	95.7	95.9	95.9	95.9	95.9	95.9
Sep-03	99.5	96.6	97.4	97.3	96.5	96.4	96.6	96.6	96.6	96.6	96.6
Oct-03	100.5	97.3	98.1	98.0	97.2	97.1	97.3	97.3	97.3	97.3	97.3
Nov-03	101.5	98.0	98.8	98.7	97.9	97.8	98.0	98.0	98.0	98.0	98.0
Dec-03	102.5	98.7	99.5	99.4	98.6	98.5	99.7	99.7	99.7	99.7	99.7
Jan-04	103.5	99.4	100.2	100.1	99.3	99.2	100.4	100.4	100.4	100.4	100.4
Feb-04	104.5	100.1	100.9	100.8	100.0	100.0	100.2	100.2	100.2	100.2	100.2
Mar-04	105.5	100.8	101.6	101.5	100.7	100.6	100.8	100.8	100.8	100.8	100.8
Apr-04	106.5	101.5	102.3	102.2	101.4	101.3	101.5	101.5	101.5	101.5	101.5
May-04	107.5	102.2	103.0	102.9	102.1	102.0	102.2	102.2	102.2	102.2	102.2
Jun-04	108.5	102.9	103.7	103.6	102.8	102.7	102.9	102.9	102.9	102.9	102.9
Jul-04	109.5	103.6	104.4	104.3	103.5	103.4	103.6	103.6	103.6	103.6	103.6
Aug-04	110.5	104.3	105.1	105.0	104.2	104.1	104.3	104.3	104.3	104.3	104.3
Sep-04	111.5	105.0	105.8	105.7	104.9	104.8	105.0	105.0	105.0	105.0	105.0
Oct-04	112.5	105.7	106.5	106.4	105.6	105.5	105.7	105.7	105.7	105.7	105.7
Nov-04	113.5	106.4	107.2	107.1	106.3	106.2	106.4	106.4	106.4	106.4	106.4
Dec-04	114.5	107.1	107.9	107.8	107.0	106.9	107.1	107.1	107.1	107.1	107.1
Jan-05	115.5	107.8	108.6	108.5	107.7	107.6	107.8	107.8	107.8	107.8	107.8
Feb-05	116.5	108.5	109.3	109.2	108.4	108.3	108.5	108.5	108.5	108.5	108.5
Mar-05	117.5	109.2	110.0	109.9	109.1	109.0	109.2	109.2	109.2	109.2	109.2
Apr-05	118.5	109.9	110.7	110.6	109.8	109.7	109.9	109.9	109.9	109.9	109.9
May-05	119.5	110.6	111.4	111.3	110.5	110.4	110.6	110.6	110.6	110.6	110.6
Jun-05	120.5	111.3	112.1	112.0	111.2	111.1	111.3	111.3	111.3	111.3	111.3
Jul-05	121.5	112.0	112.8	112.7	111.9	111.8	112.0	112.0	112.0	112.0	112.0
Aug-05	122.5	112.7	113.5	113.4	112.6	112.5	112.7	112.7	112.7	112.7	112.7
Sep-05	123.5	113.4	114.2	114.1	113.3	113.2	113.4	113.4	113.4	113.4	113.4
Oct-05	124.5	114.1	114.9	114.8	114.0	113.9	114.1	114.1	114.1	114.1	114.1
Nov-05	125.5	114.8	115.6	115.5	114.7	114.6	114.8	114.8	114.8	114.8	114.8
Dec-05	126.5	115.5	116.3	116.2	115.4	115.3	115.5	115.5	115.5	115.5	115.5
Jan-06	127.5	116.2	117.0	116.9	116.1	116.0	116.2	116.2	116.2	116.2	116.2
Feb-06	128.5	116.9	117.7	117.6	116.8	116.7	116.9	116.9	116.9	116.9	116.9
Mar-06	129.5	117.6	118.4	118.3	117.5	117.4	117.6	117.6	117.6	117.6	117.6
Apr-06	130.5	118.3	119.1	119.0	118.2	118.1	118.3	118.3	118.3	118.3	118.3
May-06	131.5	119.0	119.8	119.7	118.9	118.8	119.0	119.0	119.0	119.0	119.0
Jun-06	132.5	119.7	120.5	120.4	119.6	119.5	119.7	119.7	119.7	119.7	119.7
Jul-06	133.5	120.4	121.2	121.1	120.3	120.2	120.4	120.4	120.4	120.4	120.4
Aug-06	134.5	121.1	121.9	121.8	121.0	120.9	121.1	121.1	121.1	121.1	121.1
Sep-06	135.5	121.8	122.6	122.5	121.7	121.6	121.8	121.8	121.8	121.8	121.8
Oct-06	136.5	122.5	123.2	123.1	122.3	122.2	122.4	122.4	122.4	122.4	122.4
Nov-06	137.5	123.2	123.9	123.8	123.0	122.9	123.1	123.1	123.1	123.1	123.1
Dec-06	138.5	123.9	124.6	124.5	123.7	123.6	123.8	123.8	123.8	123.8	123.8
Jan-07	139.5	124.6	125.3	125.2	124.4	124.3	124.5	124.5	124.5	124.5	124.5
Feb-07	140.5	125.3	126.0	125.9	125.1	125.0	125.2	125.2	125.2	125.2	125.2
Mar-07	141.5	126.0	126.7	126.6	125.8	125.7	125.9	125.9	125.9	125.9	125.9
Apr-07	142.5	126.7	127.4	127.3	126.5	126.4	126.6	126.6	126.6	126.6	126.6
May-07	143.5	127.4	128.1	128.0	127.2	127.1	127.3	127.3	127.3	127.3	127.3
Jun-07	144.5	128.1	128.8	128.7	127.9	127.8	128.0	128.0	128.0	128.0	128.0
Jul-07	145.5	128.8	129.5	129.4	128.6	128.5	128.7	128.7	128.7	128.7	128.7
Aug-07	146.5	129.5	130.2	130.1	129.3	129.2	129.4	129.4	129.4	129.4	129.4
Sep-07	147.5	130.2	130.9	130.8	130.0	129.9	130.1	130.1	130.1	130.1	130.1
Oct-07	148.5	130.9	131.6	131.5	130.7	130.6	130.8	130.8	130.8	130.8	130.8
Nov-07	149.5	131.6	132.3	132.2	131.4	131.3	131.5	131.5	131.5	131.5	131.5
Dec-07	150.5	132.3	133.0	132.9	132.1	132.0	132.2	132.2	132.2	132.2	132.2
Jan-08	151.5	133.0	133.7	133.6	132.8	132.7	133.9	133.9	133.9	133.9	133.9
Feb-08	152.5	133.7	134.4	134.3	133.5	133.4	133.6	133.6	133.6	133.6	133.6
Mar-08	153.5	134.4	135.1	135.0	134.2	134.1	134.3	134.3	134.3	134.3	134.3
Apr-08	154.5	135.1	135.8	135.7	134.9	134.8	135.0	135.0	135.0	135.0	135.0
May-08	155.5	135.8	136.5	136.4	135.6	135.5	135.7	135.7	135.7	135.7	135.7
Jun-08	156.5	136.5	137.2	137.1	136.3	136.2	136.4	136.4	136.4	136.4	136.4
Jul-08	157.5	137.2	137.9	137.8	137.0	136.9	137.1	137.1	137.1	137.1	137.1
Aug-08	158.5	137.9	138.6	138.5	137.7	137.6	137.8	137.8	137.8	137.8	137.8
Sep-08	159.5	138.6	139.3	139.2	138.4	138.3	138.5	138.5	138.5	138.5	138.5
Oct-08	160.5	139.3	140.0	139.9	139.1	139.0	139.2	139.2	139.2	139.2	139.2
Nov-08	161.5	139.9	140.7	140.6	139.8	139.7	140.9	140.9	140.9	140.9	140.9
Dec-08	162.5	140.6	141.3	141.2	140.4	140.3	140.5	140.5	140.5	140.5	140.5
Jan-09	163.5	141.3	142.0	141.9	141.1	141.0	141.2	141.2	141.2	141.2	141.2
Feb-09	164.5	141.9	142.6	142.5	141.7	141.6	141.8	141.8	141.8	141.8	141.8
Mar-09	165.5	142.6	143.3	143.2	142.4	142.3	142.5	142.5	142.5	142.5	142.5
Apr-09	166.5	143.2	143.9	143.8	143.0	142.9	143.1	143.1	143.1	143.1	143.1
May-09	167.5	143.8	144.5	144.4	143.6	143.5	143.7	143.7	143.7	143.7	143.7
Jun-09	168.5	144.4	145.1	145.0	144.2	144.1	144.3	144.3	144.3	144.3	144.3
Jul-09	169.5	145.0	145.7	145.6	144.8	144.7	144.9	144.9	144.9	144.9	144.9
Aug-09	170.5	145.6	146.3	146.2	145.4	145.3	145.5	145.5	145.5	145.5	145.5
Sep-09	171.5	146.2	146.9	146.8	146.0	145.9	146.1	146.1	146.1	146.1	146.1
Oct-09	172.5	146.8	147.5	147.4	146.6	146.5	146.7	146.7	146.7	146.7	146.7
Nov-09	173.5	147.4	148.1	148.0	147.2	147.1	147.3	147.3	147.3	147.3	147.3
Dec-09	174.5	147.9	148.6	148.5	147.7	147.6	147.8	147.8	147.8	147.8	147.8
Jan-10	175.5	148.5	149.2	149.1	148.3	148.2	148.4	148.4	148.4	148.4	148.4
Feb-10	176.5	149.1	149.8	149.7	148.9	148.8	149.0	149.0	149.0	149.0	149.0
Mar-10	177.5	149.7	150.4	150.3	149.5	149.4	149.6	149.6	149.6	149.6	149.6
Apr-10	178.5	150.3	151.0	150.9	150.1	150.0	150.2	150.2	150.2	150.2	150.2
May-10	179.5	150.9	151.6	151.5	150.7	150.6	150.8	150.8	150.8	150.8	150.8
Jun-10	180.5	151.5	152.2	152.1	151.3	151.2	151.4	151.4	151.4	151.4	151.4
Jul-10	181.5	152.1	152.8	152.7	151.9	151.8	152.0	152.0	152.0	152.0	152.0
Aug-10	182.5	152.7	153.4	153.3	152.5	152.4	152.6	152.6	152.6	152.6	152.6
Sep-10	183.5	153.3	154.0	153.9	153.1	153.0	153.2	153.2	153.2	153.2	153.2
Oct-10	184.5	153.9	154.6	154.5	153.7	153.6	153.8	153.8	153.8	153.8	153.8
Nov-10	185.5	154.5	155.2	155.1	154.3	154.2	154.4	154.4	154.4	154.4	154.4
Dec-10	186.5	155.1	155.8	155.7	154.9	154.8	155.0	155.0	155.0	155.0	155.0
Jan-11	1										

Session 13 - 16 Time Series and Forecasting

Search in Presentation

Home Insert Draw Design Transitions Animations Slide Show Review View

Cut Copy New Slide Format Layout Reset Section

Font Size: 22

Font: Helvetica Neue

Text Box Picture Shapes Arrange Quick Styles Shape Outline

## Statistical Methods for Data Science

## Time Series → Which performance should be used?

- Choice of the model
- Advantages**
  - MSE gives greater weight to larger deviations (which could result from outliers)
  - MAD gives less weight to larger deviations
  - MAPE gives less overall weight to a large deviation if the time series value is large
  - LAD tells us if all deviations fall below some threshold value
- Although we illustrate all 4 techniques, in general focus will be on MSE and MAD.

## Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Shows the example for the process possible for 5 different values of  $\alpha$
  - For  $\alpha=0.7$ , it is better
  - The exponential smoothing does not have a significant initial delay due the moving average, since the initial results are not smoothed
  - Formula can be reexpressed as
- $$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

## Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Quarte	Utilization	2-Qtr	Abs(Yt - Ft)	(Yt - Ft) <sup>2</sup>	3-Qtr	Abs(Yt - Ft)	(Yt - Ft) <sup>2</sup>	4-Qtr	Abs(Yt - Ft)	(Yt - Ft) <sup>2</sup>
Jan-03	82.5									
Feb-03	81.3	81.9	0.6	0.36	81.7					
Mar-03	81.3	81.3	0	0.00	80.5	0.4	0.16			
Apr-03	79	80.2	1.15	1.32	79	1.53	2.34	81	2.03	4.12
Jan-04	76.6	77.8	1.2	1.44	77.9	2.37	5.62	79.6	2.95	8.70
Feb-04	78	77.3	0.7	0.49	77.7	0.13	0.02	78.7	0.73	0.53
Mar-04	78.4	78.2	0.2	0.04	78.1	0.73	0.53	78	0.4	0.16
Apr-04	78	78.2	0.2	0.04	78.4	0.13	0.02	77.8	0.25	0.06
Jan-05	78.8	78.4	0.4	0.16	78.5	0.4	0.16	78.3	0.5	0.25
Feb-05	78.7	78.8	0.05	0.00	78.6	0.2	0.04	78.5	0.22	0.05
Mar-05	78.4	78.6	0.15	0.02	79	0.23	0.05	78.5	0.08	0.01
Apr-05	80	79.2	0.8	0.64	79.7	0.97	0.94	79	1.02	1.04
Jan-06	80.7	80.4	0.35	0.12	80.5	1	1.00	79.5	1.25	1.56
Feb-06	80.7	80.7	0	0.00	80.7	0.23	0.05	80	0.75	0.56
Mar-06	80.8	80.8	0.05	0.00		0.07	0.00	80.6	0.25	0.06
MSE	0.33	5.85		4.64	0.84	8.39	10.94	1.42	10.4	17.11
MSE				0.33				0.84		1.43
MAD			0.42				0.65			0.87

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Moving Averages

Quarte	Utilization	2-Qtr	Abs(Yt - Ft)	(Yt - Ft) <sup>2</sup>	3-Qtr	Abs(Yt - Ft)	(Yt - Ft) <sup>2</sup>	4-Qtr	Abs(Yt - Ft)	(Yt - Ft) <sup>2</sup>
Jan-03	82.5									
Feb-03	81.3	81.9	0.6	0.36	81.7					
Mar-03	81.3	81.3	0	0.00	80.5	0.4	0.16			
Apr-03	79	80.2	1.15	1.32	79	1.53	2.34	81	2.03	4.12
Jan-04	76.6	77.8	1.2	1.44	77.9	2.37	5.62	79.6	2.95	8.70
Feb-04	78	77.3	0.7	0.49	77.7	0.13	0.02	78.7	0.73	0.53
Mar-04	78.4	78.2	0.2	0.04	78.1	0.73	0.53	78	0.4	0.16
Apr-04	78	78.2	0.2	0.04	78.4	0.13	0.02	77.8	0.25	0.06
Jan-05	78.8	78.4	0.4	0.16	78.5	0.4	0.16	78.3	0.5	0.25
Feb-05	78.7	78.8	0.05	0.00	78.6	0.2	0.04	78.5	0.22	0.05
Mar-05	78.4	78.6	0.15	0.02	79	0.23	0.05	78.5	0.08	0.01
Apr-05	80	79.2	0.8	0.64	79.7	0.97	0.94	79	1.02	1.04
Jan-06	80.7	80.4	0.35	0.12	80.5	1	1.00	79.5	1.25	1.56
Feb-06	80.7	80.7	0	0.00	80.7	0.23	0.05	80	0.75	0.56
Mar-06	80.8	80.8	0.05	0.00		0.07	0.00	80.6	0.25	0.06
MSE	0.33	5.85		4.64	0.84	8.39	10.94	1.42	10.4	17.11
MSE				0.33				0.84		1.43
MAD			0.42				0.65			0.87



# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed
- Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed
- Formula can be rearranged as

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

Quarter/ Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74

Virtual Classroom

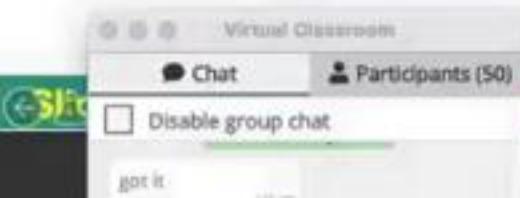
Chat

Participants (50)

Disable group chat

Sajal Ph.D (Stats)

Former Professor of Statistics | KIMS, B'lore



Sirajah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

got it

## Time Series → Smoothing Techniques – Exponential

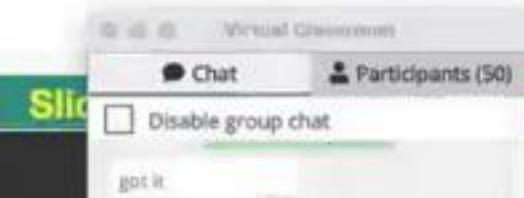
Objective is to estimate the overall time series as a combination of

- long term trend and seasonality

- Additive Model:  $X_t = F_t + S_t + \varepsilon_t$

- Multiplicative Model:  $X_t = F_t \times S_t \times \varepsilon_t$

- Multiplicative model can be converted to additive model by taking log. Other models are also possible



Time series data

an recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

	Year	ion (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$		Year	ion (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$	
Jan-03	82.5	NA	NA	NA		1	82.5	NA	NA	NA		
Feb-03	81.3	82.5	82.5	82.5		2	81.3	82.50	82.50	82.50		
Mar-03	81.3	82.26	82.02	81.66		3	81.3	82.26	82.02	81.66		
Apr-03	79	81.61	80.81	79.8		4	79	82.07	81.73	81.41		
May-04	76.6	80.61	79.13	77.56		5	76.6	81.45	80.64	79.72		
Jun-04	78	80.09	78.68	77.87		6	78	80.48	79.02	77.54		
Jul-04	78.4	78.75	78.57	78.04		7	78.4	79.99	78.61	77.86		
Aug-04						8	78	79.67	78.53	78.24		
Sep-04						9	78.8	79.34	78.32	78.07		
Oct-04						10	78.7	79.23	78.51	78.58		
Nov-04						11	78.4	79.12	78.59	78.66		
Dec-04						12	80	78.98	78.51	78.48		
Jan-05						13	80.7	79.18	79.11	79.54		
Feb-05						14	80.7	79.49	79.74	80.35		
Mar-05						15	80.8	79.73	80.13	80.60		

Exponential Smoothing

Input Range: \$B\$2:\$B\$10

Damping factor: 0.2

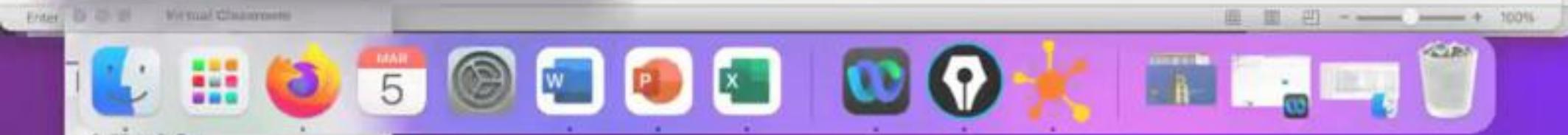
Labels

Output options

Output Range: \$C\$2:\$C\$15

Chart Output: Standard Errors

OK Cancel



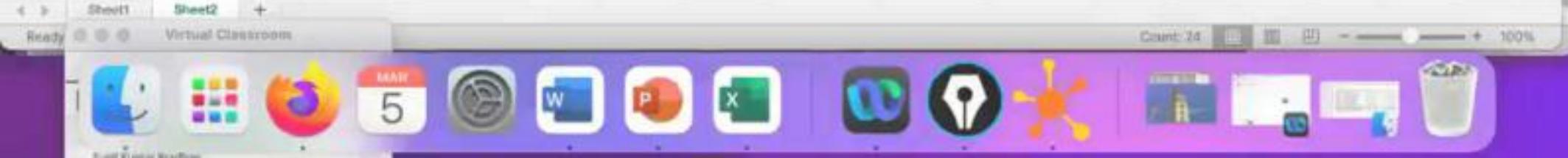
Time series data

Open recent workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

	Quarter	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$	Quarter	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
	r/year					r/year				
Jan	82.5	NA	NA	NA	1	82.5	NA	NA	NA	
Feb	81.3	82.5	82.5	82.5	2	81.3	82.50	82.50	82.50	
Mar	81.3	82.26	82.02	81.66	3	81.3	82.26	82.02	81.66	81/A
Apr	79	81.61	80.81	79.8	4	79	82.07	81.73	81.41	82.3
Jan-04	76.6	80.61	79.13	77.56	5	76.6	81.45	80.64	79.72	81.54
Feb-04	78	80.09	78.68	77.87	6	78	80.48	79.02	77.54	81.348
Mar-04	78.4	79.75	78.57	78.24	7	78.4	79.99	78.61	77.86	79.4698
Apr-04	78	79.4	78.34	78.07	8	78	79.67	78.53	78.24	77.1739
Jan-05	78.8	79.28	78.52	78.58	9	78.8	79.34	78.32	78.07	77.8347
Feb-05	78.7	79.16	78.59	78.66	10	78.7	79.23	78.51	78.58	78.286956
Mar-05	78.4	79.01	78.52	78.48	11	78.4	79.12	78.59	78.66	78.057391
Apr-05	80	79.21	79.11	79.54	12	80	78.98	78.51	78.48	78.651478
Jan-06	80.7	79.51	79.75	80.35	13	80.7	79.18	79.11	79.54	78.890295
Feb-06	80.7	79.75	80.13	80.6	14	80.7	79.49	79.74	80.35	78.458859
Mar-06	80.8	79.96	80.4	80.74	15	80.8	79.73	80.13	80.60	79.601411

The chart displays the utilization values from the table against time. A blue line represents an exponential trend fit to the data. The x-axis shows months from Jan to Mar, and the y-axis shows utilization values from 73 to 83.



Home Insert Draw Design Unavailable Accessibility Help

Stat Time Series

Methods for Data Science

Smooth Techniques - Exponential

Time Series

Smooth Techniques - Exponential

Time Series

Methods for Data Science

Smooth Techniques - Exponential

Time Series

Statistical Methods for Data Science

Time Series → Decomposition view

observed

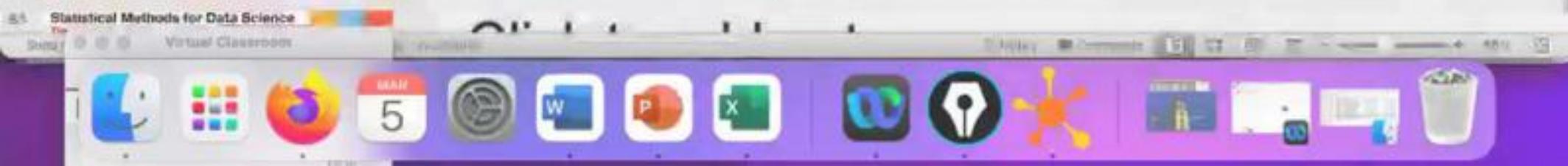
trend

seasonal

(noise)

Slide 82 | Prof.Gangaboralah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore



Session 13 – 14 Time Series and Forecasting

# Statistical Methods for Data Science

## Time Series → Smoothing Techniques – Exponential

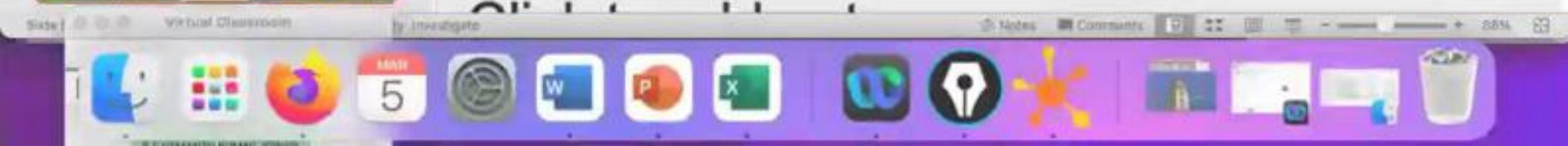
- Done the analysis on the previous problem for 3 different values of  $\alpha$ .
- For  $\alpha=0.7$ , it is better.
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial  $k$  values are not smoothed.
- Formula can be rearranged as

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

Quarter/Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
4/2006	-	79.94	80.40	80.74

Slide 88 | Prof.Gangaboralah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore



Session 13 – 14 Time Series and Forecasting

# Statistical Methods for Data Science

## Time Series → Definition of Stochastic Process

In stochastic time series,

$$Y_t, t \in \mathbb{Z} = \{\pm 1, \pm 2, \dots\}$$

is a family of random variables,  $Y_t$  denoting the value of the characteristic of interest at time  $t$ .

Thus,  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  is seen as a realized value of the random vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$  with joint probability density function  $f_{\mathbf{Y}}(\mathbf{y})$ .

Methods for Data Science  
Definition of Stochastic Process

- A stochastic process is a collection of family of random variables  $(X_t, t \in T)$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ .
- Sample space
- $\mathcal{F}$  – all collection of subsets A of  $\Omega$  and  $P$  – probability measure
- Index set

Methods for Data Science  
Definition of Stochastic Process

Index set  $T$

- The index set  $T$  is a collection of all time functions that can result from random experiment, usually the index  $T$  denote time.
- $X_t$  are independent and identically distributed (iid) random variables with mean  $\mu_t$  and variance  $\sigma_t^2$ . Each realization of this process gives an ensemble or a data set.

Statistical Methods for Data Science  
Time Series → Definition of Stochastic Process

In stochastic time series,

$$Y_t, t \in \mathbb{Z} = \{\pm 1, \pm 2, \dots\}$$

is a family of random variables,  $Y_t$  denoting the value of the characteristic of interest at time  $t$ .

Thus,  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  is seen as a realized value of the random vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$  with joint probability density function  $f_{\mathbf{Y}}(\mathbf{y})$ .

Statistical Methods for Data Science  
Time Series → Definition of Stochastic Process

Virtual Classroom

By: Investigate



Session 13 - 14 Time Series and Forecasting

Home Insert Draw Design Transitions Animations Slide Show Review View

Layout Reset Section Helvetica Neue 22 Font Size

Picture Shapes Text Box Arrange Quick Styles Shape Outline

Methods for Data Science Definition of Stochastic Process

- A stochastic process is a collection of family random variables  $\{X_t, t \in T\}$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ .
- Sample space
- $\mathcal{F}$  - all collection of subsets A of  $\Omega$  and  $P$  - probability measure
- Index set

Methods for Data Science Definition of Stochastic Process

Index set  $T$ :  
The index set  $T$  is a collection of all time functions that can result from random experiment, usually the index  $T$  denote time.  $X_t$  are independent and identically distributed (iid) random variables with mean  $\mu_t$  and variance  $\sigma_t^2$ . Each realization of this process gives an ensemble or a data set.

Statistical Methods for Data Science Time Series Definition of Stochastic Process

In stochastic time series,  
 $Y_t, t \in \mathbb{Z} = \{\pm 1, \pm 2, \dots\}$   
is a family of random variables,  $Y_t$  denoting the value of the characteristic of interest at time  $t$ . Thus,  $y = (y_1, y_2, \dots, y_n)^T$  is seen as a realized value of the random vector  $Y = (Y_1, Y_2, \dots, Y_n)^T$  with joint probability density function  $f_Y(y)$ .

# Statistical Methods for Data Science

## Time Series → Definition of Stochastic Process

### Index set:

- The index set  $T$  is a collection of all time functions that can result from random experiment, usually the index  $T$  denote time.  $X_t$  are independent and identically distributed (iid) random variables with mean  $\mu_t$  and variance  $\sigma_t^2$ . Each realization of this process gives an ensemble or a data set.

Slide 85 | Prof.Gangaboralah PhD (Stats) | Former Professor of Statistics | KIMS, Bangalore

Virtual Classroom By: Investigate Notes Comments

5 MAR

Icons: Phone, Grid, Firefox, Calendar, Clock, Word, Powerpoint, Excel, WPS Office, Notepad, Brain, Document, Folder, Recycle Bin

Session 13 - 14 Time Series and Forecasting

# Statistical Methods for Data Science

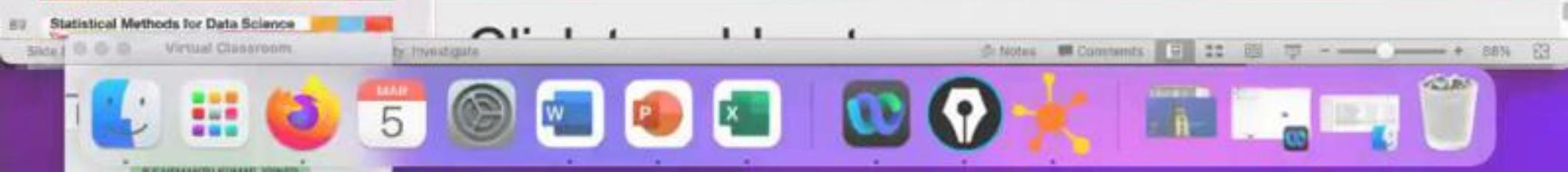
## Time Series → Definition of Stochastic Process

The joint distribution function of a finite random variables

$$\{Y_{t_1}, \dots, Y_{t_n}\}, t_1 < t_2 < \dots < t_n$$

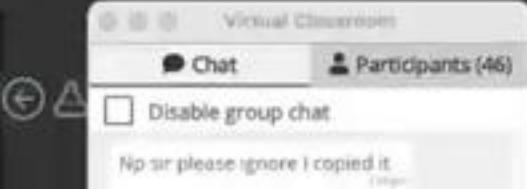
from the collection  $\{Y_t, t \in T\}$  is

$$F_{Y_{t_1}, \dots, Y_{t_n}}(y_1, \dots, y_n) = P[Y_{t_1} \leq y_{t_1}, \dots, P[Y_{t_n} \leq y_{t_n}],$$

$$(y_{t_1}, \dots, y_{t_n}) \in R^n$$


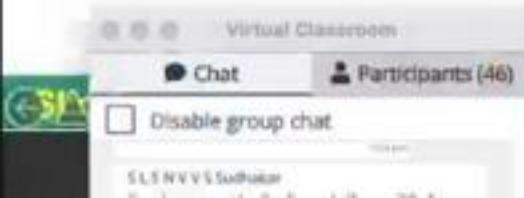
# Static

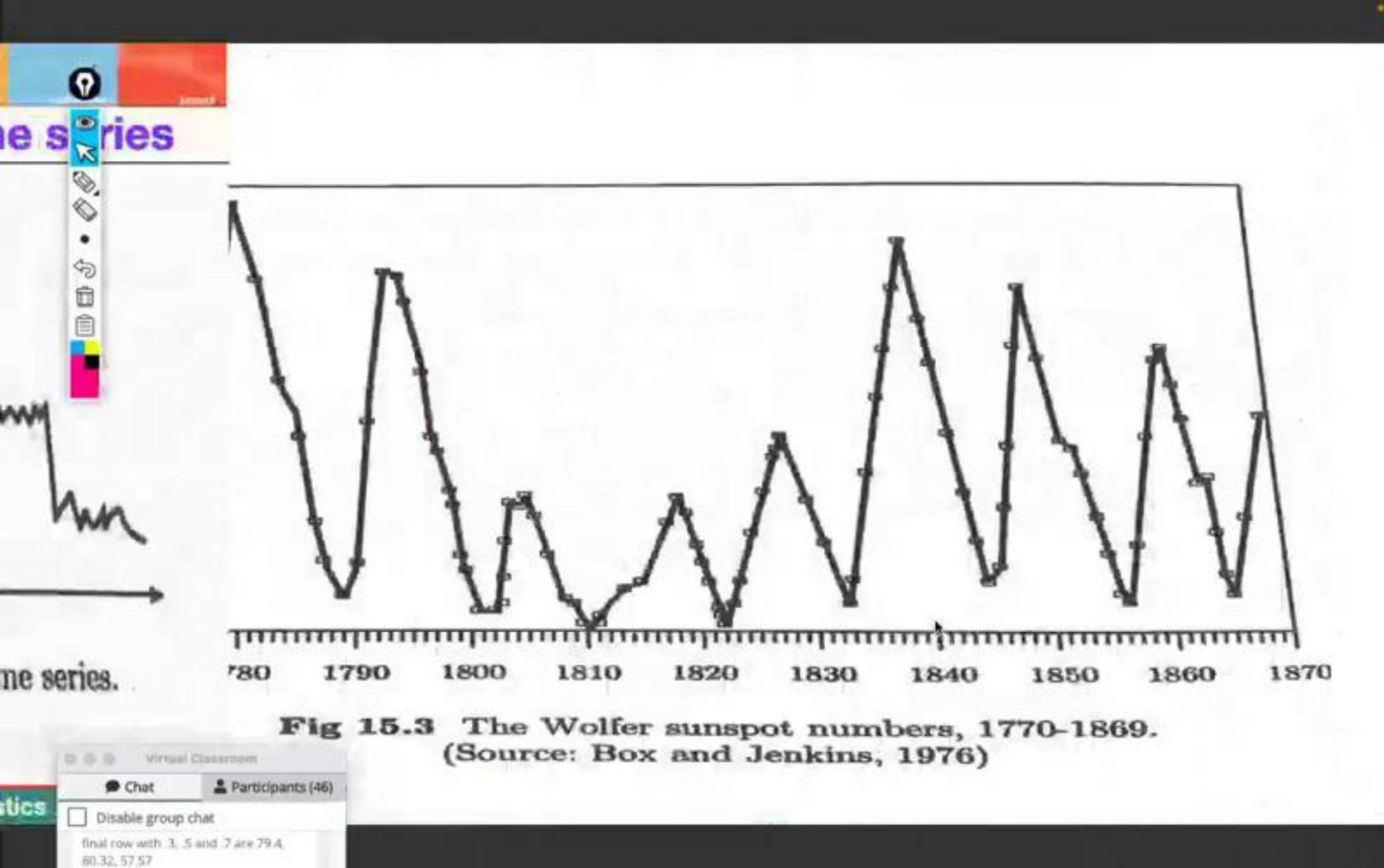




: A special kind of stochastic process is based on the assumption that the process is in a particular state of equilibrium. This type of assumption is called stationarity.

A stochastic process is strictly stationary if its properties are unaffected by a change of origin.



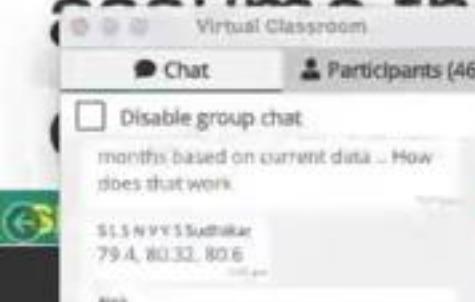


**Fig 15.3** The Wolfer sunspot numbers, 1770-1869.  
(Source: Box and Jenkins, 1976)

## Time Series → Stationary Stochastic Process in Time series

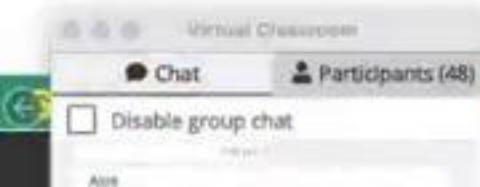
: However, many non-stationary time series can be so modified that the reduced to time series obeys the original series.

The two main component which cause lack of stationarity are trend and seasonality. In fitting the stationary time series model, we therefore, ~~remove~~ <sup>assume</sup> that the trend and seasonality have been from the original series.



- Stationarity implies that the joint distribution of pair of observations  $y_{t_1}, y_{t_2}$  viz.,  
$$f(y_{t_1}, y_{t_2}) = f(y_{t_1+h}, y_{t_2+h})$$

For any integer  $h$ . That is, the joint distribution of any pair of observations on time points which differ by a constant quantity is the same for such pairs.



# Science

ress



## Autocovariance

level



diagrams for lag  $k$  ( $k = 1, 2$ ).

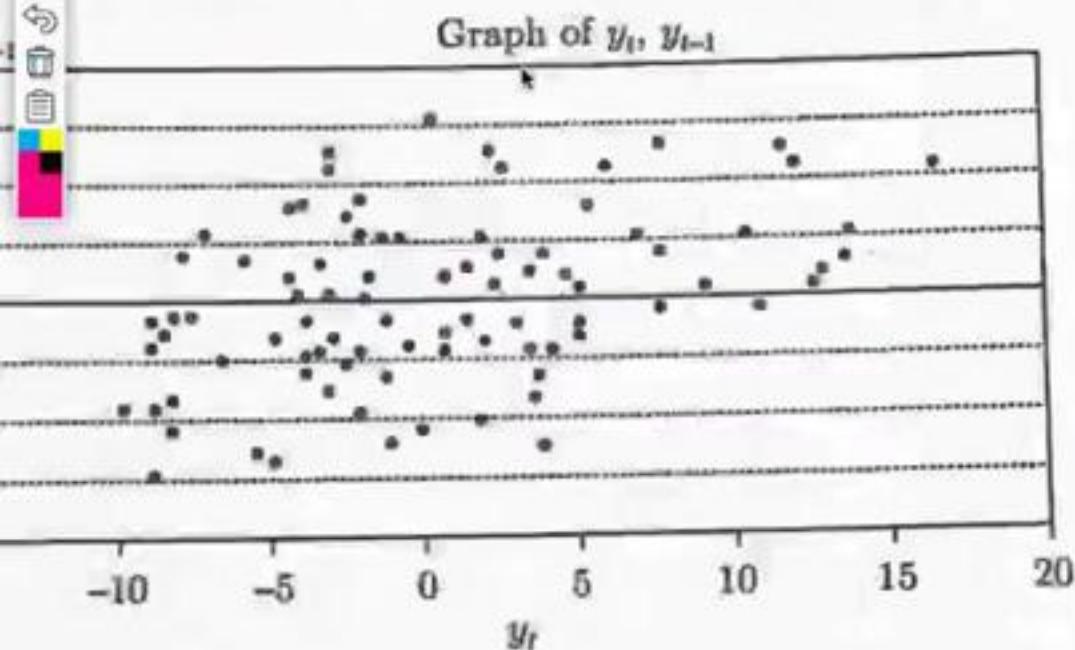


Fig 15.4(b)  $Y_t = 0.6e_{t-1} + e_t$

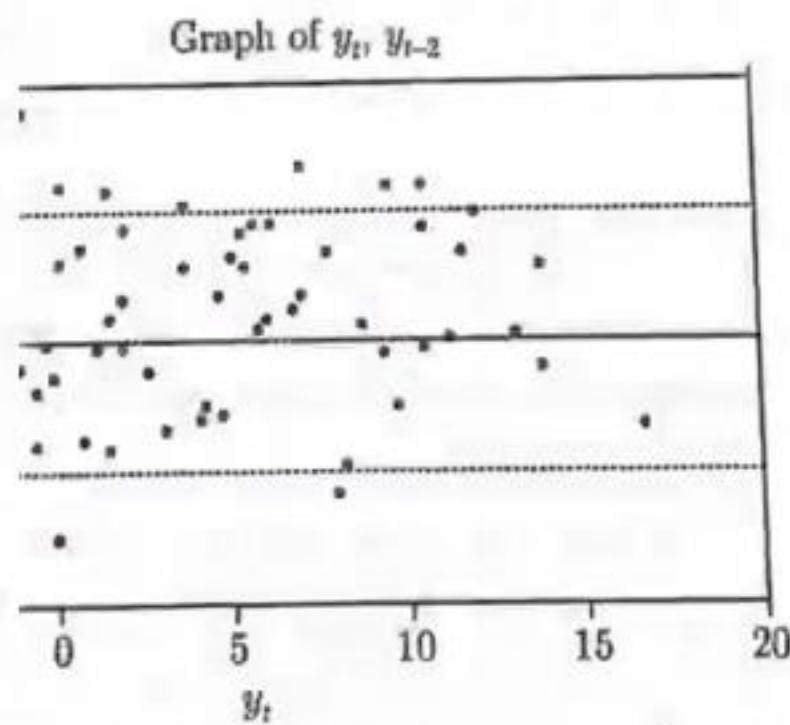


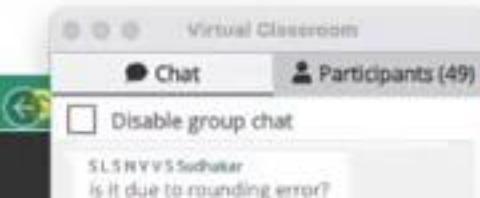
Fig 15.5(b)  $Y_t = 0.7e_t + e_{t-1}$

Virtual Classroom

- Chat
- Participants (48)
- Disable group chat
- Brute force approach?

## Time Series → Autocovariance and Autocorrelation

- Let  $\{Y_t\}$  be a Time Series with  $E\{Y^2\} < \infty$ . The mean function of  $\{Y_t\}$  is  $\mu_y(t) = E(Y_t)$ .
- The covariance function of  $\{Y_t, Y_{t+h}\}$  is
- $\gamma_h = \text{Cov}(Y_t, Y_{t+h}) = E[(Y_t - \mu)(Y_{t+h} - \mu)]$



Sriram Ph.D (Stats)

Former Professor of Statistics | KIMS, Bangalore

# Statistical Methods for Data Science

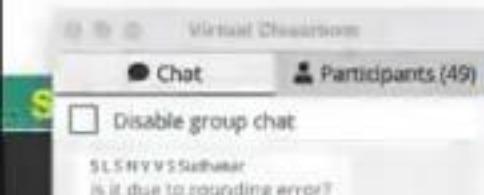
## Time Series → Autocovariance and Autocorrelation

- Let  $\{Y_t\}$  be a Time Series with  $E\{Y^2\} < \infty$ . The mean function of  $\{Y_t\}$  is  $\mu_y(t) = E(Y_t)$ .
- The covariance function of  $\{Y_t, Y_{t+h}\}$  is
- $\gamma_h = \text{Cov}(Y_t, Y_{t+h}) = E[(Y_t - \mu)(Y_{t+h} - \mu)]$



- Clearly,

$$|\gamma_h| \leq \gamma_0, \forall h = 1, 2, \dots$$



Soraiah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore



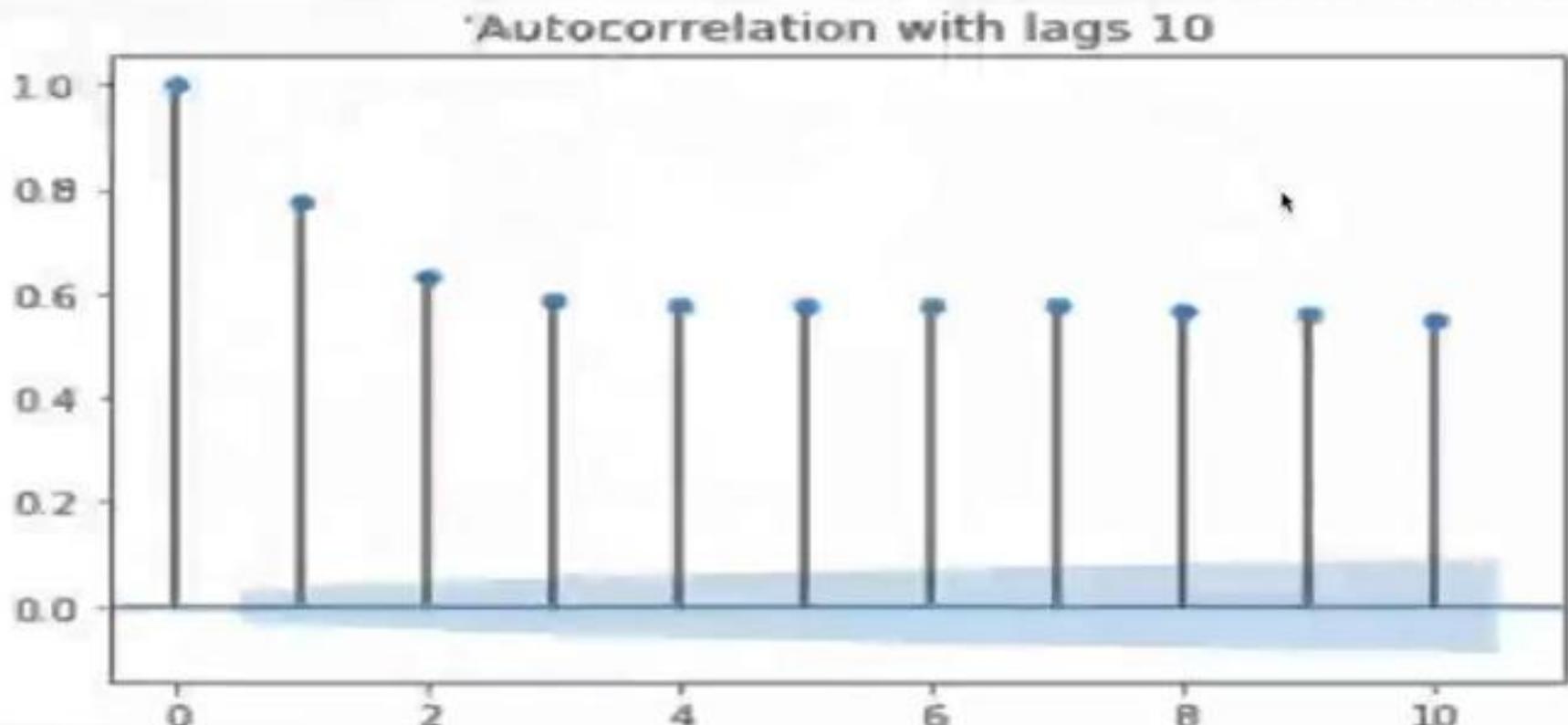
O-  
SS

Virtual Classroom

Chat    Participants (48)

Disable group chat

SLS N V V S Sudhaier  
Is it due to rounding error?



Virtual Classroom

Chat Participants (48)

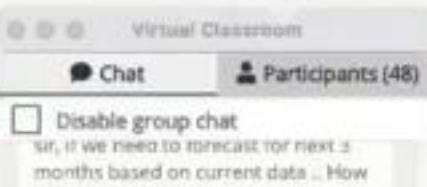
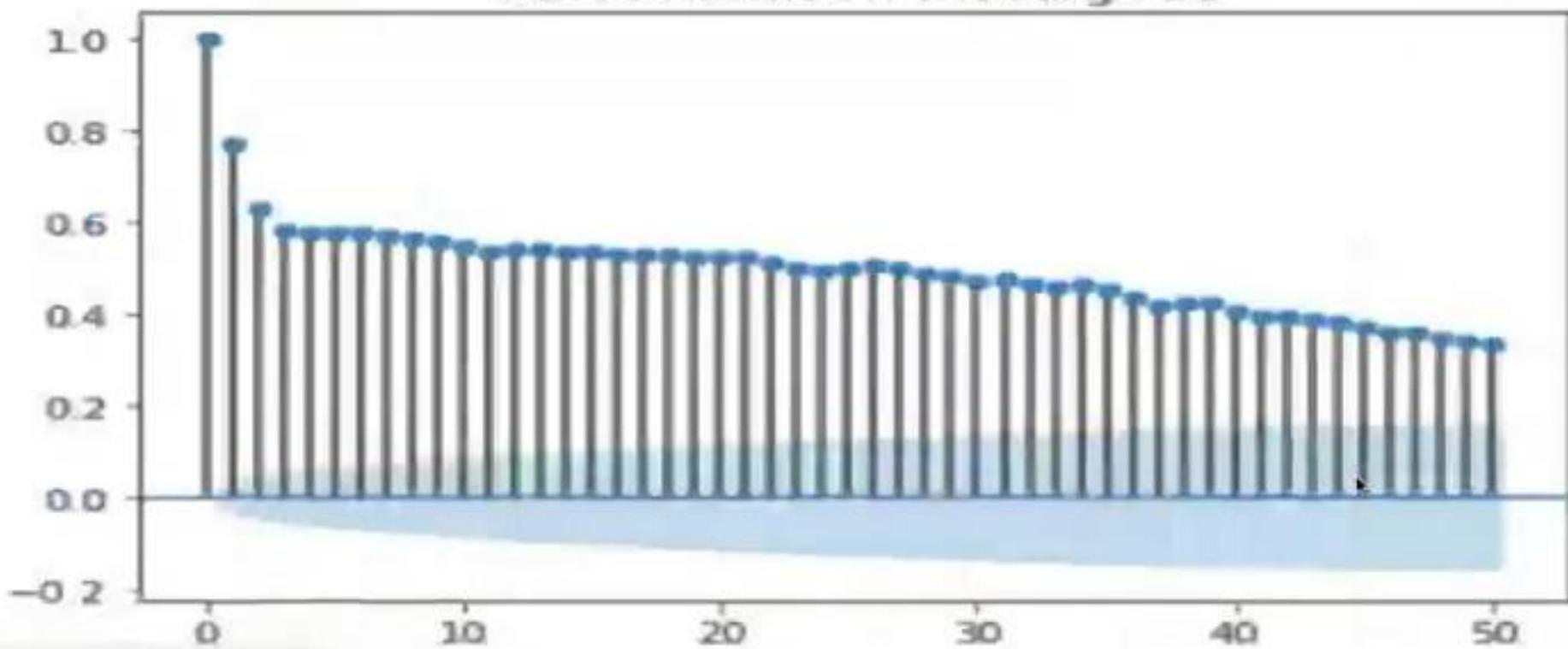
Disable group chat

sir, if we need to forecast for next 3 months based on current data ... How do we start

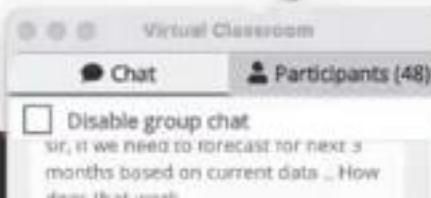
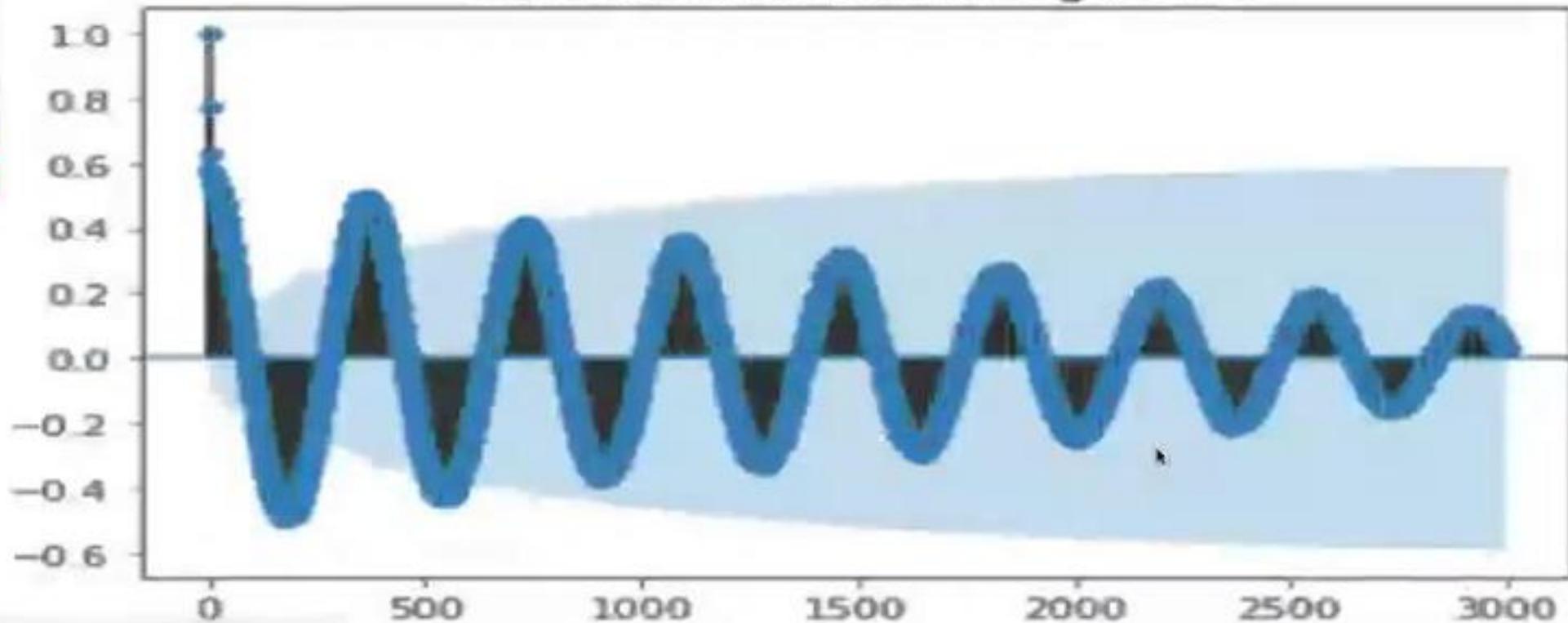
Sorajah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

### 'Autocorrelation with lags 50



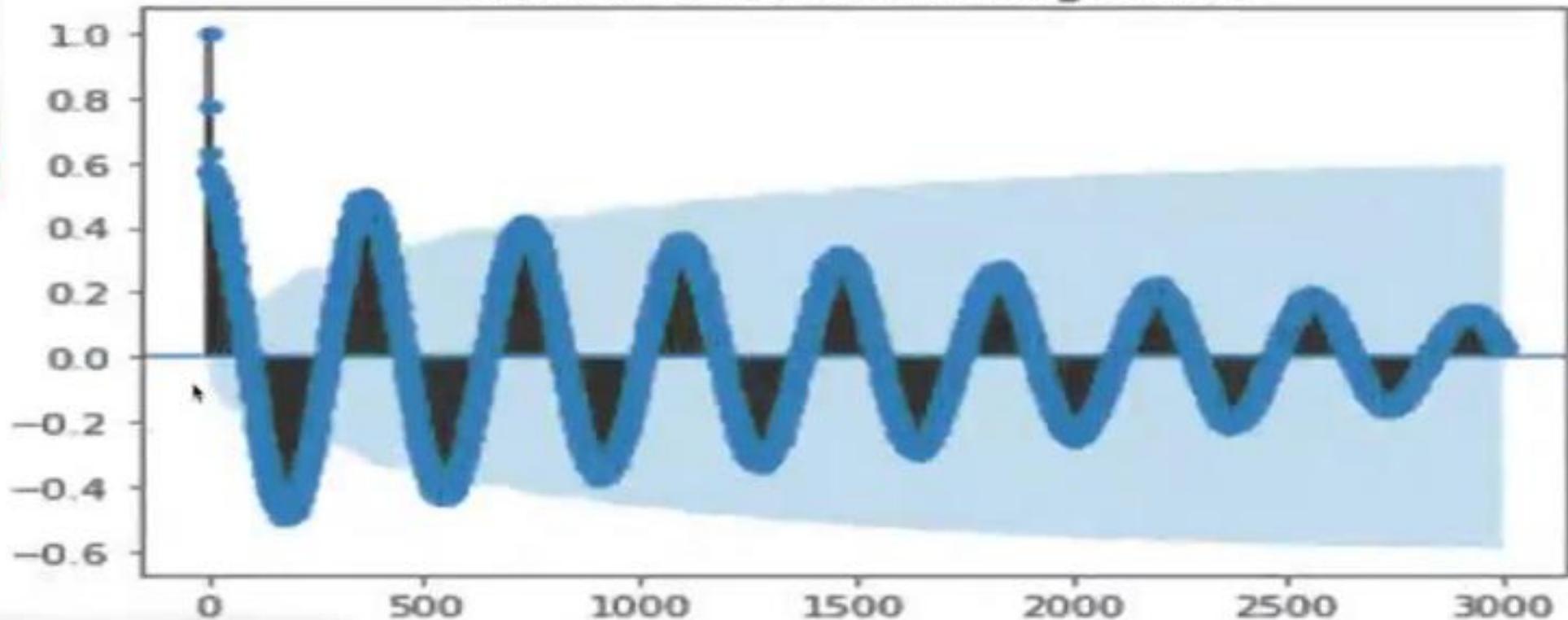
### 'Autocorrelation with lags 3000



## Time Series → Autocorrelation



Autocorrelation with lags 3000



Virtual Classroom

Chat Participants (48)

Disable group chat

Sur, if we need to forecast for next 3 months based on current data ... How does that work?

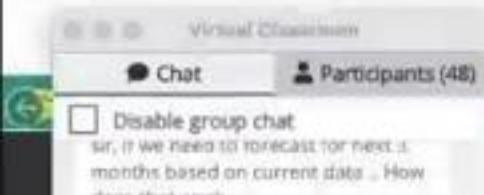
Sorajah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore

## Time Series → Partial - Autocorrelation Function (PACF)

- The partial correlation between two variables is a conditional correlation taking into account their dependence on all other remaining variables
- Equation for partial autocorrelation function is

$$\frac{\text{Cov}(X_t, X_{t-3} | X_{t-1}, X_{t-2})}{\sqrt{\text{Var}(X_t | X_{t-1}, X_{t-2}) \text{Var}(X_{t-3} | X_{t-1}, X_{t-2})}}$$



Sorajah PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

# Statistical Methods for Data Science

Time Series

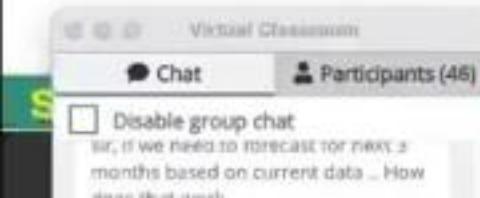
→ Partial – Autocorrelation Function (PACF)

- The partial correlation between two variables is a conditional correlation taking into account their dependence on all other remaining variables
- Equation for partial autocorrelation function is

$$\frac{\text{Cov}(X_t, X_{t-3} | X_{t-1}, X_{t-2})}{\sqrt{\text{Var}(X_t | X_{t-1}, X_{t-2}) \text{Var}(X_{t-3} | X_{t-1}, X_{t-2})}}$$



- To try to determine which one of these forecasting methods gives the “best” forecast, we evaluate the “success” of each method using a *performance measure*



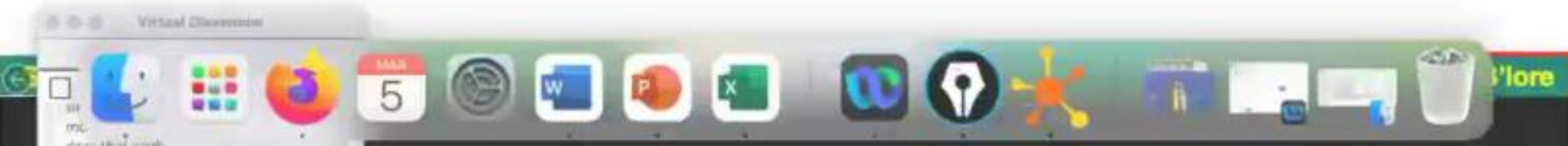
Sorah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore

# Statistical Methods for Data Science

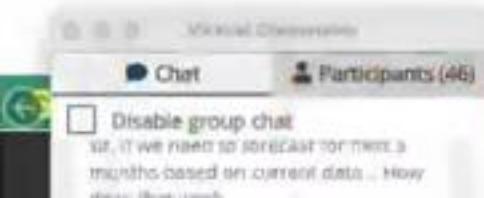
## Time Series → Measurement of forecast error

- To try to determine which one of these forecasting methods gives the “best” forecast, we evaluate the “success” of each method using a ***performance measure***



## Time Series → Measurement of forecast error

- To try to determine which one of these forecasting methods gives the “best” forecast, we evaluate the “success” of each method using a ***performance measure***



Soraiyah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore

## Time Series → Which performance should be used?

- Choice of the modeler

- **Advantages**

- MSE gives greater weight to larger deviations (which could result from outliers)



Sathakar PhD (Stats)

Former Professor of Statistics | KIMS, Bangalore

The screenshot shows a video conference interface. On the left, there is a vertical toolbar with various icons for navigation and control. In the center, a large video frame displays a presentation slide with text and graphics. To the right of the video frame is a "Virtual Classroom" window. This window has tabs for "Chat" and "Participants (45)". The "Chat" tab is active, showing a list of messages from participants. The messages include:

- Alok: And identifying the best alpha is the brute force approach?
- SLS N V S Sudhakar: Is it due to rounding error?
- Gunasekaran R: Hera which is the actual value
- Alok: 5 mins break sir?
- Sal Kumar K C: Sir, Do we get problems in ARMA and ARIMA Model in the exam? or it is only theory questions?
- K Lakshmi Prasanna: Sir 5 mins break?
- Nagaraj Ganesh Nayak: can you please speak more on conditional correlation used in PACF?
- Alok: Is break started sir?

At the bottom of the chat window, there is a message input field with placeholder text "writing to everyone" and a "Write a message" button.

At the very bottom of the screen, there is a dock with icons for various applications, including Microsoft Office (Word, Excel, PowerPoint), a drawing tool, and a trash bin.



Chat

Participants (44)

 Disable group chat

SLSN VVS Sudhakar  
final row with .3,.5 and .7 are 79.4,  
80.32, 57.57

Alok  
Sir, if we need to forecast for next 3  
months based on current data .. How  
does that work?

SLSN VVS Sudhakar  
79.4, 80.32, 80.6

Alok  
And identifying the best alpha is the  
brute force approach ?

SLSN VVS Sudhakar  
Is it due to rounding error?

Ganesan R  
Here which is the actual value

Alok  
5 mins break sir ?

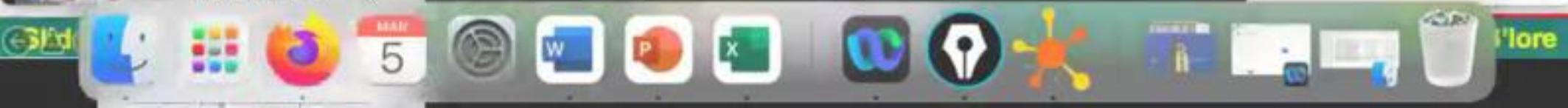
Sai Kumar KC  
Sir, Do we get problems in ARMA and  
ARIMA Model in the exam? or it is only  
theory questions?

K Lamine Prasanna  
Sir 5 mins break?

Nagaraj Ganesh Nagre  
can you please speak more on  
conditional correlation used in PACF?

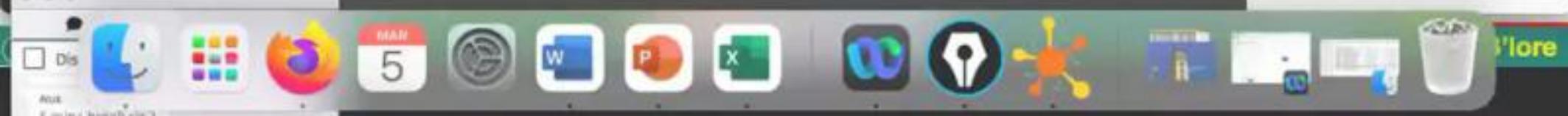
Alok  
is break started sir ?

Sending to everyone  
Write Message





Chat Participants (43)

 Disable group chatSLS NVV S Sudhakar  
79.4, 80.32, 80.6Akash  
And identifying the best alpha is the brute force approach?SLS NVV S Sudhakar  
is it due to rounding error?Guna Sekaran R  
Hera which is the actual valueAkash  
5 mins break sir?Sai Kumar K C  
Sir, Do we get problems in ARMA and ARIMA Model in the exam? or it is only theory questions?K Lakshmi Prasanna  
Sir 5 mins break?Naganaj Ganesh Nayak  
Can you please speak more on conditional correlation used in PACF?Akash  
is break started sir?Suril Kumar Pradhan  
yes sirMukul Mathur  
ues sirAbhishek Pandey  
next saturday will be fineSending to everyone  
Write Message > |

## Time Series → Smoothing Techniques – Exponential

- Done the analysis on the previous problem for 3 different values of  $\alpha$

Quarter/ Year	Utilization (%)	$F_t(\alpha=0.2)$	$F_t(\alpha=0.4)$	$F_t(\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.26	82.02	81.66
4/2003	79.00	81.61	80.81	79.80
1/2004	76.60	80.61	79.13	77.56
2/2004	78.00	80.09	78.68	77.87
3/2004	78.40	79.75	78.57	78.24
4/2004	78.00	79.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.70	79.16	78.59	78.66
3/2005	78.40	79.01	78.52	78.48
4/2005	80.00	79.21	79.11	79.54
1/2006	80.70	79.51	79.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.96	80.40	80.74
MSE				

Virtual Classroom

 Chat Participants (44) Disable group chat

ARIMA Model in the exam? or it is only theory questions?

Abdullah Ph.D (Stats)

Former Professor of Statistics | KIMS, Bangalore

**Statistical Methods for Data Science**

**Time Series → Smoothing Techniques – Exponential**

- Done the analysis on the previous problem for 3 different values of  $\alpha$
- For  $\alpha=0.7$ , it is better
- The exponential smoothing does not have a significant initial delay like the Moving Av. where the initial values are not smoothed
- Formula can be rearranged as

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

Quarter Year	Utilization (%)	$F_t (\alpha=0.2)$	$F_t (\alpha=0.4)$	$F_t (\alpha=0.7)$
1/2003	82.50			
2/2003	81.30	82.50	82.50	82.50
3/2003	81.30	82.20	82.02	81.68
4/2003	79.00	81.51	80.81	79.89
1/2004	78.80	80.61	79.13	77.58
2/2004	78.00	80.08	78.68	77.87
3/2004	78.40	78.75	78.57	78.24
4/2004	78.00	78.40	78.34	78.07
1/2005	78.80	79.28	78.52	78.58
2/2005	78.20	78.15	78.50	78.68
3/2005	78.40	78.01	78.52	78.48
4/2005	78.00	79.21	79.11	79.54
1/2006	80.70	79.81	78.75	80.35
2/2006	80.70	79.75	80.13	80.60
3/2006	80.80	79.98	80.40	80.74
MSE				

Excel File Edit View Insert Formulas Tools Data Window Help Prof.Gangaborsah B Andanalah PhD Sat 5 Mar 7:59 PM

Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Calibri (Body) 20 A A+ Wrap Text Number Conditional Formatting

Format as Table Cell Styles Insert Delete Format

Auto-sum Filter Sort & Filter Find & Select

Save recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

File Save As Open Recent Undo Redo Cut Copy Paste

Cells fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	79	81.61	80.81	79.8		4	79	82.07	81.73	81.41			9.41	7.46	5.80	
2	76.6	80.61	79.13	77.56		5	76.6	81.45	80.64	79.72			23.57	16.32	9.75	
3	78	80.09	78.68	77.87		6	78	80.48	79.02	77.54			6.17	1.05	0.21	
4	78.4	79.75	78.57	78.24		7	78.4	79.99	78.61	77.86			2.52	0.05	0.29	
5	78	79.4	78.34	78.07		8	78	79.67	78.53	78.24			2.79	0.28	0.06	
6	78.8	79.28	78.52	78.58		9	78.8	79.34	78.32	78.07			0.29	0.23	0.53	
7	78.7	79.16	78.59	78.66		10	78.7	79.23	78.51	78.58			0.28	0.04	0.01	
8	78.4	79.01	78.52	78.48		11	78.4	79.12	78.59	78.66			0.52	0.03	0.07	
9	80	79.21	79.11	79.54		12	80	78.98	78.51	78.48			1.04	2.22	2.31	
10	80.7	79.51	79.75	80.35		13	80.7	79.18	79.11	79.54			2.30	2.54	1.34	
11	80.7	79.75	80.13	80.6		14	80.7	79.49	79.74	80.35			1.47	0.91	0.12	
12	80.8	79.96	80.4	80.74		15	80.8	79.73	80.13	80.60			1.15	0.45	0.04	
13						16		79.94	80.40	80.74						
14													68.56	39.8		
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																

Sheet1 Sheet2 +

Microsoft Classroom

Dis Firefox Calendar W P X WPS Office Word Excel OneDrive

Excel File View Insert Format Tools Data Window Help Prof. Gangaborsiah B Andanaiyah PhD Sat 5 Mar 8:00 PM

Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Calibri (Body) 20 A A+ Wrap Text Number Conditional Formatting Themes as Table Cell Styles Insert Delete Format Sort & Filter Find & Select

Recent workbooks? Your recent changes were saved. Do you want to continue working where you left off?

fx =H3-F3)^2

Utilizat

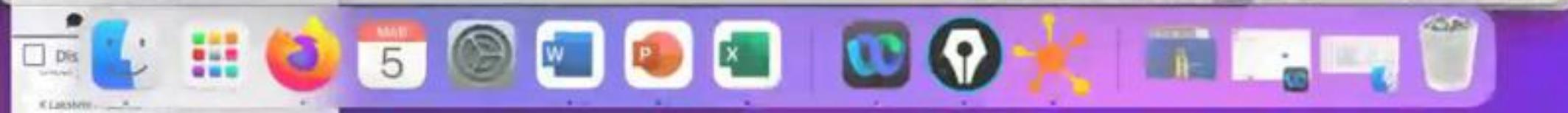
n (%) F<sub>t</sub>(a=0.2) F<sub>t</sub>(a=0.4) F<sub>t</sub>(a=0.7)

Quarte Utilizat

r/year ion (%) F<sub>t</sub>(a=0.2) F<sub>t</sub>(a=0.4) F<sub>t</sub>(a=0.7)

MSE (0.2) MSE (0.4) MSE (0.7)

1	82.5	NA	NA	NA	1	82.5	NA	NA	NA	NA	NA	NA	NA
2	81.3	82.5	82.5	82.5	2	81.3	82.50	82.50	82.50	NA	1.44	NA	NA
3	81.3	82.26	82.02	81.66	3	81.3	82.26	82.02	81.66	NA	0.92	0.52	0.13
4	79	81.61	80.81	79.8	4	79	82.07	81.73	81.41	NA	9.41	7.46	5.80
5	76.6	80.61	79.13	77.56	5	76.6	81.45	80.64	79.72	NA	23.57	16.32	9.75
6	78	80.09	78.68	77.87	6	78	80.48	79.02	77.54	NA	6.17	1.05	0.21
7	78.4	79.75	78.57	78.24	7	78.4	79.99	78.61	77.86	NA	2.52	0.05	0.29
8	78	79.4	78.34	78.07	8	78	79.67	78.53	78.24	NA	2.79	0.28	0.06
9	78.8	79.28	78.52	78.58	9	78.8	79.34	78.32	78.07	NA	0.29	0.23	0.53
10	78.7	79.16	78.59	78.66	10	78.7	79.23	78.51	78.58	NA	0.28	0.04	0.01
11	78.4	79.01	78.52	78.48	11	78.4	79.12	78.59	78.66	NA	0.52	0.03	0.07
12	80	79.21	79.11	79.54	12	80	78.98	78.51	78.48	NA	1.04	2.22	2.31
13	80.7	79.51	79.75	80.35	13	80.7	79.18	79.11	79.54	NA	2.30	2.54	1.34
14	80.7	79.75	80.13	80.6	14	80.7	79.49	79.74	80.35	NA	1.47	0.91	0.12
15	80.8	79.96	80.4	80.74	15	80.8	79.73	80.13	80.60	NA	1.15	0.45	0.04



Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboralah B Andanalah PhD Sat 6 Mar 8:00 PM

Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Calibri (Body) A A+ A- Wrap Text Number Conditional Formatting Cell Styles Insert Delete Format Auto-sum Sort & Filter Find & Select

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off? Yes No

f(x) =SUM[M3:M16]

B	C	D	E	F	G	H	I	J	K	L	M	N	O
76.6	80.61	79.13	77.56		5	76.6	81.45	80.64	79.72		23.57	16.32	9.75
78	80.09	78.68	77.87		6	78	80.48	79.02	77.54		6.17	1.05	0.21
78.4	79.75	78.57	78.24		7	78.4	79.99	78.61	77.86		2.52	0.05	0.29
78	79.4	78.34	78.07		8	78	79.67	78.53	78.24		2.79	0.28	0.06
78.8	79.28	78.52	78.58		9	78.8	79.34	78.32	78.07		0.29	0.23	0.53
78.7	79.16	78.59	78.66		10	78.7	79.23	78.51	78.58		0.28	0.04	0.01
78.4	79.01	78.52	78.48		11	78.4	79.12	78.59	78.66		0.52	0.03	0.07
80	79.21	79.11	79.54		12	80	78.98	78.51	78.48		1.04	2.22	2.31
80.7	79.51	79.75	80.35		13	80.7	79.18	79.11	79.54		2.30	2.54	1.34
80.7	79.75	80.13	80.6		14	80.7	79.49	79.74	80.35		1.47	0.91	0.12
80.8	79.96	80.4	80.74		15	80.8	79.73	80.13	80.60		1.15	0.45	0.04
					16		79.94	80.40	80.74		53.87		
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													

Sheet Sheet2 + Virtual Classroom

100%



Timeseries data

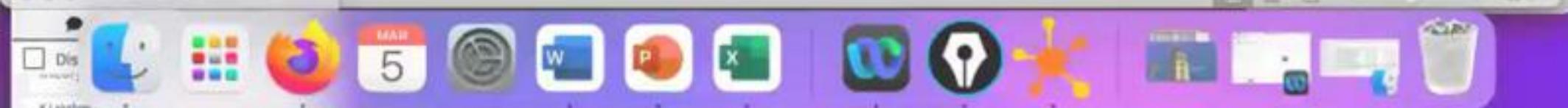
Do you want to save recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
●	76.6	80.61	79.13	77.56		5	76.6	81.45	80.64	79.72		23.57	16.32	9.75	
◀	78	80.09	78.68	77.87		6	78	80.48	79.02	77.54		6.17	1.05	0.21	
●	78.4	79.75	78.57	78.24		7	78.4	79.99	78.61	77.86		2.52	0.05	0.29	
●	78	79.4	78.34	78.07		8	78	79.67	78.53	78.24		2.79	0.28	0.06	
●	78.8	79.28	78.52	78.58		9	78.8	79.34	78.32	78.07		0.29	0.23	0.53	
●	78.7	79.16	78.59	78.66		10	78.7	79.23	78.51	78.58		0.28	0.04	0.01	
●	78.4	79.01	78.52	78.48		11	78.4	79.12	78.59	78.66		0.52	0.03	0.07	
●	80	79.21	79.11	79.54		12	80	78.98	78.51	78.48		1.04	2.22	2.31	
●	80.7	79.51	79.75	80.35		13	80.7	79.18	79.11	79.54		2.30	2.54	1.34	
●	80.7	79.75	80.13	80.6		14	80.7	79.49	79.74	80.35		1.47	0.91	0.12	
●	80.8	79.96	80.4	80.74		15	80.8	79.73	80.13	80.60		1.15	0.45	0.04	
						16		79.94	80.40	80.74		53.87	33.53	22.11	

17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															

Sheet1 Sheet2 + Virtual Classroom



Time series data

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	82.5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
2	81.3		82.50	82.50	82.50			1.44		1.44		1.44								
3	81.3		82.26	82.02	81.66			0.92		0.52		0.13								
4	79		82.07	81.73	81.41			9.41		7.46		5.80								
5	76.6		81.45	80.64	79.72			23.57		16.32		9.75								
6	78		80.48	79.02	77.54			6.17		1.05		0.21								
7	78.4		79.99	78.61	77.86			2.52		0.05		0.29								
8	78		79.67	78.53	78.24			2.79		0.28		0.06								
9	78.8		79.34	78.32	78.07			0.29		0.23		0.53								
10	78.7		79.23	78.51	78.58			0.28		0.04		0.01								
11	78.4		79.12	78.59	78.66			0.52		0.03		0.07								
12	80		78.98	78.51	78.48			1.04		2.22		2.31								
13	80.7		79.18	79.11	79.54			2.30		2.54		1.34								
14	80.7		79.49	79.74	80.35			1.47		0.91		0.12								
15	80.8		79.73	80.13	80.60			1.15		0.45		0.04								
16			79.94	80.40	80.74			53.87		33.53		22.11								
17																				
18																				

MSE

3.591235366

2.2355817

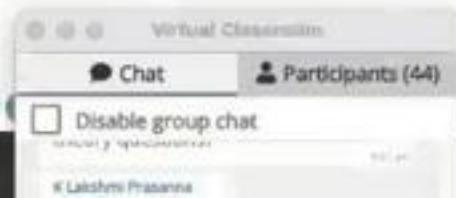
1.473684





- An **autoregressive (AR) model** is a representation of some type of random process; used to describe certain time-varying processes. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic (random) error term; thus the model is in the form of a stochastic difference equation.


$$y_t = c_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$



Turnitin detection

Home Insert Draw Design Layout Formulas Data Page Number Cell Conditional Formatting Format as Table Spell Check Find & Select

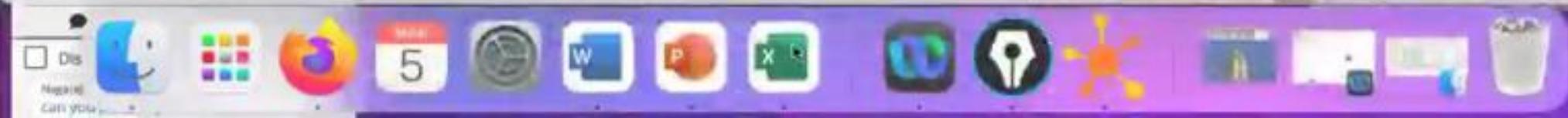
Cut (Shift + C) Copy (Shift + C) Paste (Shift + V) Undo (Shift + Z) Redo (Shift + Y) Find & Select

Format Cells (Shift + F1)

Recent worksheets: WORK RECENT CHARTS, WORKBOOK1, L10 (you have to continue working in more you will edit)

=M17/25

	D2,J	NA	NA	NA	NA	NA	NA	NA	NA
2	81.3	82.50	82.50	82.50		1.44	1.44	1.44	
3	81.3	82.26	82.02	81.66		0.92	0.52	0.13	
4	79	82.07	81.73	81.41		9.41	7.46	5.80	
5	76.6	81.45	80.64	79.72		23.57	16.32	9.75	
6	78	80.48	79.02	77.54		6.17	1.05	0.21	
7	78.4	79.99	78.61	77.86		2.52	0.05	0.29	
8	78	79.67	78.53	78.24		2.79	0.28	0.06	
9	78.8	79.34	78.32	78.07		0.29	0.23	0.53	
10	78.7	79.23	78.51	78.58		0.28	0.04	0.01	
11	78.4	79.12	78.59	78.66		0.52	0.03	0.07	
12	80	78.98	78.51	78.48		1.04	2.22	2.31	
13	80.7	79.18	79.11	79.54		2.30	2.54	1.34	
14	80.7	79.49	79.74	80.35		1.47	0.91	0.12	
15	80.8	79.73	80.13	80.60		1.15	0.45	0.04	
16		79.94	80.40	80.74		53.87	33.53	22.11	
	MSE					3.59	2.34	1.47	



Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboralaiah B Andanaiah PhD Sat 5 Mar 8:04 PM

Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Helvetica Ne... 22 A A+ Wrap Text General Conditional Formatting Format as Table Cell Styles Insert Delete Format Auto-sum Fill Clear Sort & Filter Find & Select

ben recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

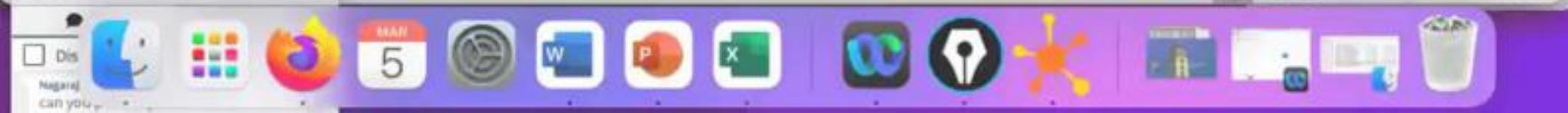
Uarte Utilizat

year	ion (%)	$F_t(a=0.2)$	$F_t(a=0.4)$	$F_t(a=0.7)$	MSE (0.2)	MSE (0.4)	MSE
1	82.5	NA	NA	NA	NA	NA	NA
2	81.3	82.50	82.50	82.50		1.44	1.44
3	81.3	82.26	82.02	81.66		0.92	0.52
4	79	82.07	81.73	81.41		9.41	7.46
5	76.6	81.45	80.64	79.72		23.57	16.32
6	78	80.48	79.02	77.54		6.17	1.05
7	78.4	79.99	78.61	77.86		2.52	0.05
8	78	79.67	78.53	78.24		2.79	0.28
9	78.8	79.34	78.32	78.07		0.29	0.23
10	78.7	79.23	78.51	78.58		0.28	0.04
11	78.4	79.12	78.59	78.66		0.52	0.03
12	80	78.98	78.51	78.48		1.04	2.22
13	80.7	79.18	79.11	79.54		2.30	2.54
14	80.7	79.49	79.74	80.35		1.47	0.91
15	80.8	79.73	80.13	80.60		1.15	0.45

Sheet1 Sheet2 +

Virtual Classroom

100%



Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Calibri (Body) t1 A A= = Wrap Text General Conditional Formatting Cell Styles Insert Delete Format Auto-Sum Sort & Filter Find & Select

sen recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

fx =VARP(H17)

	H	I	J	K	L	M	N	O	P	Q	R	S
1	82.5	81.3						NA	NA	NA		NA
2	81.3	79						82.50	82.50	82.50		
3	81.3	76.6						82.26	82.02	81.66		
4	79	78						82.07	81.73	81.41		
5	76.6	78.4						81.45	80.64	79.72		
6	78	78						80.48	79.02	77.54		
7	78.4	78.8						79.99	78.61	77.86		
8	78	78.7						79.67	78.53	78.24		
9	78.8	78.4						79.34	78.32	78.07		
10	78.7	80						79.23	78.51	78.58		
11	78.4	80.7						79.12	78.59	78.66		
12	80	80.7						78.98	78.51	78.48		
13	80.7	80.8						79.18	79.11	79.54		
14	80.7							79.49	79.74	80.35		
15	80.8							79.73	80.13	80.60		
Mean	79.55	=VARP(H17)						79.94	80.40	80.74		

Formula Builder

Show All Functions VARP Number1 = 79.54666667 H17 Number2 = number

+ Result: 0

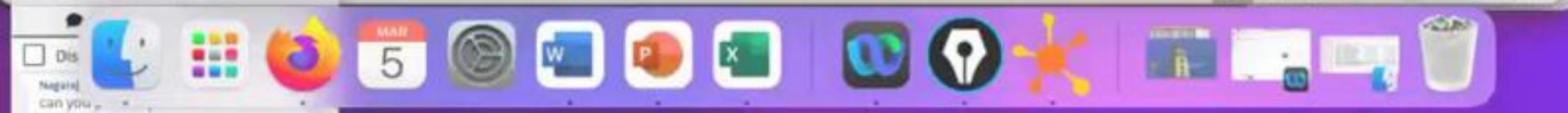
fx VARP

This function is available for compatibility with Excel 2007 and earlier.

Calculates variance based on the entire population (ignores logical values and text in the population).

More help on this function

Sheet1 Sheet2 + Virtual Classroom



Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Calibri (Body) 22 A A = Wrap Text Number Conditional Formatting Auto-sum

Format Text Box Merge & Centre % Insert Delete Format

Format as Table Cell Styles Sort & Filter Find & Select

Save recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

	H	I	J	K	L	M	N	O	P	Q	R	S
4	79	78							82.07	81.73	81.41	
5	76.6	78.4							81.45	80.64	79.72	
6	78	78							80.48	79.02	77.54	
7	78.4	78.8							79.99	78.61	77.86	
8	78	78.7							79.67	78.53	78.24	
9	78.8	78.4							79.34	78.32	78.07	
10	78.7	80							79.23	78.51	78.58	
11	78.4	80.7							79.12	78.59	78.66	
12	80	80.7							78.98	78.51	78.48	
13	80.7	80.8	+						79.18	79.11	79.54	
14	80.7								79.49	79.74	80.35	
15	80.8								79.73	80.13	80.60	
Mean	79.55	2.47							79.94	80.40	80.74	MSE

Formula Builder

Show All Functions

VARP

Number1 = {82.07;81.73;81.41;79.72;80.48;79.99;78.61;77.86;79.67;78.53;78.24;79.34;78.32;78.07;79.23;78.51;78.58;79.12;78.59;78.66;78.98;78.51;78.48;79.18;79.11;79.54;79.49;79.74;80.35;79.73;80.13;80.60}

Number2 = number

Result: 2.489155566 Done

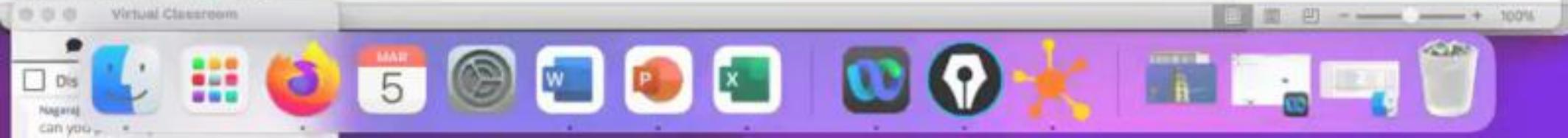
f x VARP

This function is available for compatibility with Excel 2007 and earlier.

Calculates variance based on the entire population (ignores logical values and text in the population).

Quotient

More help on this function



Time series data

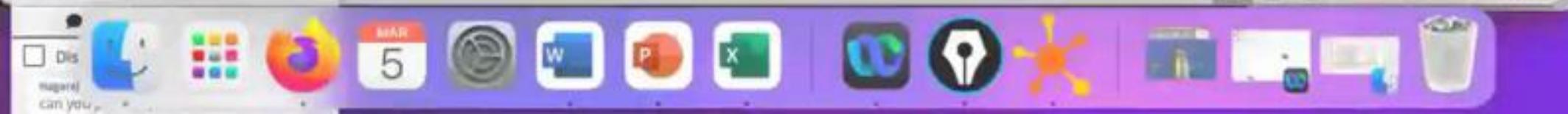
sen recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Quarter/ year	Utilization (%)	lag 2	y <sub>t</sub> =y <sub>t-1</sub> -mean	+/-	F <sub>t</sub> ( $\alpha=0.2$ )	F <sub>t</sub> ( $\alpha=0.4$ )	F <sub>t</sub> ( $\alpha=0.7$ )	MSE (0.2)	MSE (0.4)	MSE (0.7)
1	82.5	81.3	2.95		NA	NA	NA	NA	NA	NA
2	81.3	79	1.75		82.50	82.50	82.50	1.44	1.44	
3	81.3	76.6	1.75		82.26	82.02	81.66	0.92	0.52	
4	79	78	-0.55		82.07	81.73	81.41	9.41	7.46	
5	76.6	78.4	-2.95		81.45	80.64	79.72	23.57	16.32	
6	78	78	-1.55		80.48	79.02	77.54	6.17	1.05	
7	78.4	78.8	-1.15		79.99	78.61	77.86	2.52	0.05	
8	78	78.7	-1.55		79.67	78.53	78.24	2.79	0.28	
9	78.8	78.4	-0.75		79.34	78.32	78.07	0.29	0.23	
10	78.7	80	-0.85		79.23	78.51	78.58	0.28	0.04	
11	78.4	80.7	-1.15		79.12	78.59	78.66	0.52	0.03	
12	80	80.7	0.45		78.98	78.51	78.48	1.04	2.22	
13	80.7	80.8	1.15		79.18	79.11	79.54	2.30	2.54	
14	80.7				79.49	79.74	80.35	1.47	0.91	
15	80.8				79.73	80.13	80.60	1.15	0.45	

Sheet1 Sheet2 +

Virtual Classroom

100%

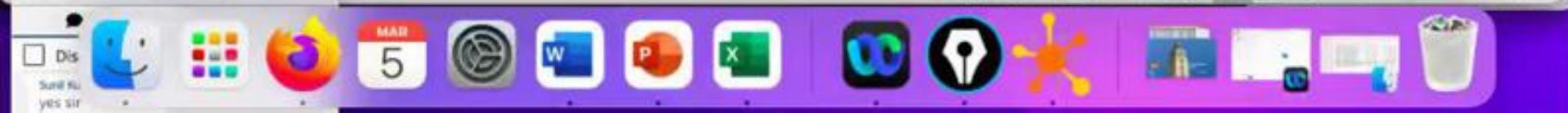


Time series data

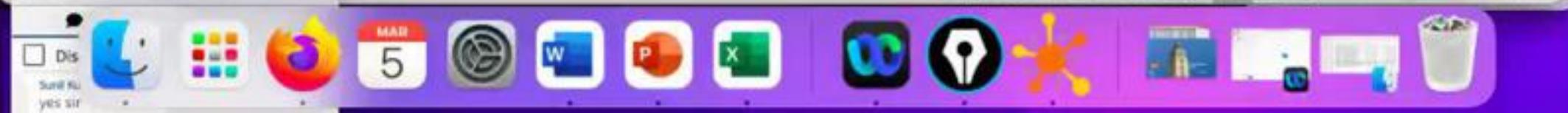
Has changes been recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
8	78	78.7	-1.55	-0.85	1.3095					79.67	78.53	78.24			2.79
9	78.8	78.4	-0.75	-1.15	0.8562					79.34	78.32	78.07			0.29
10	78.7	80	-0.85	0.45	-0.384					79.23	78.51	78.58			0.28
11	78.4	80.7	-1.15	1.15	-1.322					79.12	78.59	78.66			0.52
12	80	80.7	0.45	1.15	0.5228					78.98	78.51	78.48			1.04
13	80.7	80.8	1.15	1.25	1.4455					79.18	79.11	79.54			2.30
14	80.7				8.9536					79.49	79.74	80.35			1.47
15	80.8				Autocovariance 0.6887					79.73	80.13	80.60			1.15
16	Mean	79.55			Autocorrelation					79.94	80.40	80.74			53.87
17	Variance	2.47											MSE		3.59

Time series data											Prof.Gangaboraiyah B Andanalal PhD		Sat 5 Mar 8:13 PM		
Home Insert Draw Page Layout Formulas Data Review View											Search Sheet		Shares		
Cut	Calibri (Body)	22	A	A	=	=	Wrap Text	General	Auto-sum	Auto-sum	Sum	Sum	Sum	Sum	
Paste	Copy	B	I	U			Merge & Centre	Conditional	Format as Table	Cell Styles	Insert	Delete	Format	Clear	
Format	Format							Formatting	Format					Sort & Filter	
Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?															
C23	x	v	fx												
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
	The following data gives the cargo exports $y_t$ (in units of $10^3$ metric tons) at the principal port in India for a period of 20 years														
	Find autocovariance, autocorrelation and auto regression														
Year (t)	$y_t$	lag 1 ( $y_1 = y_{t-1}$ )	lag 2 ( $y_2 = y_{t-2}$ )	lag3 ( $y_3 = y_{t-3}$ )	lag 1 ( $y_4 = y_{t-4}$ )	lag 5 ( $y_5 = y_{t-5}$ )	$y_{t-m}$	$y_{1-m}$	$y_{2-m}$	$y_{3-m}$	$y_{4-m}$	$y_{5-m}$	$(y_{t-m})$	$(y_{t-m})$	
1	1230	1345	1382	1416	1593	1802	-1010.2	-895	-858	-824	-647	-438.2	9E+05	9E+	
2	1345	1382	1416	1593	1802	1817	-895.2	-858	-824	-647	-438	-423.2	8E+05	7E+	
3	1382	1416	1593	1802	1817	1995	-858.2	-824	-647	-438	-423	-243	-245.2	7E+05	6E+
4	1416	1593	1802	1817	1995	2212	-824.2	-647	-438	-423	-245	-28.2	5E+05	4E+	
5	1593	1802	1817	1995	2212	2607	-647.2	-438	-423	-245	-28	366.8	3E+05	3E+	
6	1802	1817	1995	2212	2607	2698	-438.2	-423	-245	-28	367	457.8	2E+05	1E+	
7	1817	1995	2212	2607	2698	2703	-423.2	-245	-28	367	458	462.8	1E+05	119	
8	1995	2212	2607	2698	2703	2738	-245.2	-28	367	458	463	497.8	6915	-899	
9	2212	2607	2698	2703	2738	2810	-28.2	367	458	463	498	569.8	-10344	-129	
10	2607	2698	2703	2738	2810	2810	366.8	458	463	498	570	391.8	2E+05	2E+	
11	2698	2703	2738	2810	2810	2632	457.8	463	498	570	392	307.8	2E+05	2E+	
12	2703	2738	2810	2632	2548	2881	462.8	498	570	392	308	640.8	2E+05	3E+	
13	2738	2810	2632	2548	2881	2725	497.8	570	392	308	641	484.8	3E+05	2E+	
14	2810	2632	2548	2881	2725	2911	569.8	392	308	641	485	670.8	2E+05	2E+	
15	2632	2548	2881	2725	2911	2759	391.8	308	641	485	671	518.8	1F+05	3F+	



Time series data												Prof.Gangaboraiyah B Andanalal PhD	Sat 5 Mar 8:13 PM				
Home	Insert	Draw	Page Layout	Formulas	Data	Review	View									Q Search Sheet	Share
Cut	Calibri (Body)	22	A A	= =	= =	Wrap Text	General	Auto-sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum
Paste	Copy	B I U				Merge & Centre	Conditional Formatting	Format as Table	Cell Styles	Insert	Delete	Format	Clear	Sort & Filter	Find & Select		
Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?																Yes	
C23	x	v	f(x)													No	
Principal port in India for a period of 20 years																	
lag 5 ( $y_t = y_{t-5}$ )	$y_{t-m}$	$y_{t-m}$	$y_{t-m}$	$y_{t-m}$	$y_{t-m}$	$y_{t-m}$	$y_{t-m}$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$	$(y_{t-m})(y_{t-m})$		
1802	-1010.2	-895	-858	-824	-647	-438.2	9E+05	9E+05	832607	653801	442670						
1817	-895.2	-858	-824	-647	-438	-423.2	8E+05	7E+05	579373	392277	378849						
1995	-858.2	-824	-647	-438	-423	-245.2	7E+05	6E+05	376063	363190	210431						
2212	-824.2	-647	-438	-423	-245	-28.2	5E+05	4E+05	348801	202094	23242						
2607	-647.2	-438	-423	-245	-28	366.8	3E+05	3E+05	158693	18251	-2E+05						
2698	-438.2	-423	-245	-28	367	457.8	2E+05	1E+05	12357	-2E+05	-2E+05						
2703	-423.2	-245	-28	367	458	462.8	1E+05	11934	-2E+05	-2E+05	-2E+05						
2738	-245.2	-28	367	458	463	497.8	6915	-89939	-1E+05	-1E+05	-1E+05						
2810	-28.2	367	458	463	498	569.8	-10344	-12910	-13051	-14038	-16068						
2632	366.8	458	463	498	570	391.8	2E+05	2E+05	182593	209003	143712						
2548	457.8	463	498	570	392	307.8	2E+05	2E+05	260854	179366	140911						
2881	462.8	498	570	392	308	640.8	2E+05	3E+05	181325	142450	296562						
2725	497.8	570	392	308	641	484.8	3E+05	2E+05	153223	318990	241333						
2911	569.8	392	308	641	485	670.8	2E+05	2E+05	365128	276239	382222						
2759	391.8	308	641	485	671	518.2	1F+05	3P+05	189945	767819	709966						

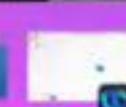
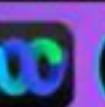
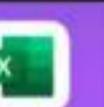


Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes  No

principal port in India for a period of 20 years

lag 5 ( $y_{t-5}$ )	$y_{t-m}$	$y_{1-m}$	$y_{2-m}$	$y_{3-m}$	$y_{4-m}$	$y_{5-m}$	$(y_{t-m})$	$(y_{t-m})^2$	$(y_{t-m})^3$	$(y_{t-m})^4$	$(y_{t-m})^5$
1802	-1010.2	-895	-858	-824	-647	-438.2	9E+05	9E+05	832607	653801	442670
1817	-895.2	-858	-824	-647	-438	-423.2	8E+05	7E+05	579373	392277	378849
1995	-858.2	-824	-647	-438	-423	-245.2	7E+05	6E+05	376063	363190	210431
2212	-824.2	-647	-438	-423	-245	-28.2	5E+05	4E+05	348801	202094	23242
2607	-647.2	-438	-423	-245	-28	366.8	3E+05	3E+05	158693	18251	-2E+05
2698	-438.2	-423	-245	-28	367	457.8	2E+05	1E+05	12357	-2E+05	-2E+05
2703	-423.2	-245	-28	367	458	462.8	1E+05	11934	-2E+05	-2E+05	-2E+05
2738	-245.2	-28	367	458	463	497.8	6915	-89939	-1E+05	-1E+05	-1E+05
2810	-28.2	367	458	463	498	569.8	-10344	-12910	-13051	-14038	-16068
2632	366.8	458	463	498	570	391.8	2E+05	2E+05	182593	209003	143712
2548	457.8	463	498	570	392	307.8	2E+05	2E+05	260854	179366	140911
2881	462.8	498	570	392	308	640.8	2E+05	3E+05	181325	142450	296562
2725	497.8	570	392	308	641	484.8	3E+05	2E+05	153223	318990	241333
2911	569.8	392	308	641	485	670.8	2E+05	2E+05	365128	276239	382222
2759	391.8	308	641	485	671	518.8	1E+05	3E+05	189945	262819	203266



Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboraiyah B Andanalal PhD Sat 5 Mar 8:13 PM

Time series data

Home Insert Draw Page Layout Formulas Data Review View

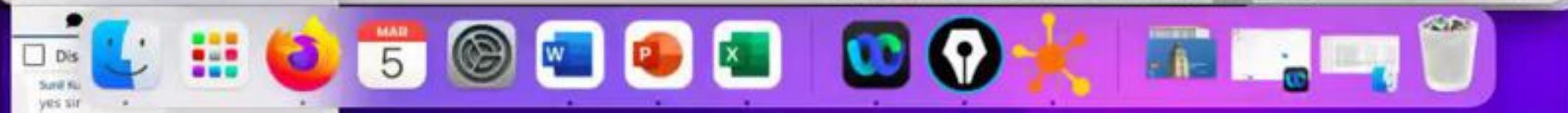
Cut Copy Format Calibri (Body) 22 A A Wrap Text General Conditional Formatting Cell Styles Auto-sum Sort & Filter

Paste B I U Merge & Centre % Insert Delete Format Clear

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off? Yes No

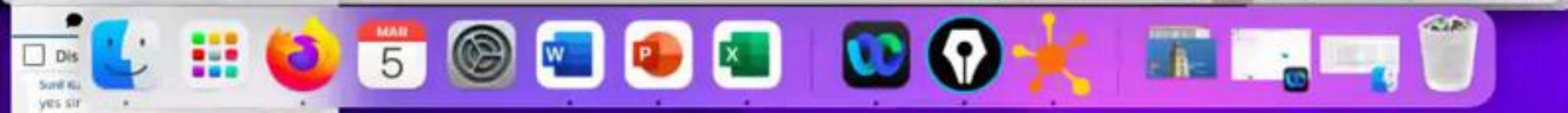
principal port in India for a period of 20 years

lag 5 ( $y_{t-5}$ )	$y_{t-m}$	$y_{1-m}$	$y_{2-m}$	$y_{3-m}$	$y_{4-m}$	$y_{5-m}$	$(y_{t-m})(y_{1-m})$	$(y_{t-m})(y_{2-m})$	$(y_{t-m})(y_{3-m})$
1802	-1010.2		-895.2	-858.2	-824.2	-647.2	-438.2	904331.04	866953.64
1817	-895.2		-858.2	-824.2	-647.2	-438.2	-423.2	768260.64	737823.84
1995	-858.2		-824.2	-647.2	-438.2	-423.2	-245.2	707328.44	555427.04
2212	-824.2		-647.2	-438.2	-423.2	-245.2	-28.2	533422.24	361164.44
2607	-647.2		-438.2	-423.2	-245.2	-28.2	366.8	283603.04	273895.04
2698	-438.2		-423.2	-245.2	-28.2	366.8	457.8	185446.24	107446.64
2703	-423.2		-245.2	-28.2	366.8	457.8	462.8	103768.64	11934.24
2738	-245.2		-28.2	366.8	457.8	462.8	497.8	6914.64	-89939.36
2810	-28.2		366.8	457.8	462.8	497.8	569.8	-10343.76	-12909.96
2632	366.8		457.8	462.8	497.8	569.8	391.8	167921.04	169755.04
2548	457.8		462.8	497.8	569.8	391.8	307.8	211869.84	227892.84
2881	462.8		497.8	569.8	391.8	307.8	640.8	230381.84	263703.44
2725	497.8		569.8	391.8	307.8	640.8	484.8	283646.44	195038.04
2911	569.8		391.8	307.8	640.8	484.8	670.8	223247.64	175384.44
2759	391.8		307.8	640.8	484.8	670.8	518.8	120596.04	251065.44



Time series data											Prof.Gangaboraiyah B Andanalal PhD	Sat 5 Mar 8:13 PM	
Home Insert Draw Page Layout Formulas Data Review View											Search Sheet	Share	
Cut	Calibri (Body)	22	A-	A+	=	Wrap Text	General	Conditional Formatting	Format as Table	Cell Styles	Auto-sum	Sort & Filter	Find & Select
Copy	B	I	U			Merge & Centre					Sum		
Paste											Fill		
Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?													
11	L	M	N	O	P	Q	R	S	T	U	V	W	
18	670.8	518.8	120596.04	251065.44	189944.64	262819.44	203265.84						
19	518.8		197238.24	149221.44	206472.24	159686.64							
20			310659.84	429848.64	332447.04								
21			325203.84	251514.24									
22			348011.04										
23													
24													
25	Count		19	18	17	16	15						
26	Sumproduct	5901506.96	4925219.12	3899349.48	2696177.84	1691211							
27	autocovariance	310605.6295	273623.2844	229373.4988	168511.115	112747.4							
28	Autocorrelation	0.909211544	0.80095602	0.671427087	0.493269395	0.330036638							
29													
30													
31													
32													
33													
34													
35													
	Sheet1	Sheet2	+										
	Virtual Classroom												
	Dis	Call	Firefox	Calendar	Word	PowerPoint	Excel	OneNote	Skype	OneDrive	Office	Recycle Bin	

Time series data											Q Search Sheet	Shares			
Home	Insert	Draw	Page Layout	Formulas	Data	Review	View								
Paste	Cut	Copy	Format	Calibri (Body)	22	A A+	= =	Wrap Text	Number	Conditional Formatting	Format as Table	Cell Styles	Insert	Delete	Format
Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?															
P28	X	f1	+P27/B26												
Find autocovariance, autocorrelation and auto regression															
Year (t)	y <sub>t</sub>	lag 1 (y <sub>1</sub> = y <sub>t-1</sub> )	lag 2 (y <sub>2</sub> = y <sub>t-2</sub> )	lag3 (y <sub>3</sub> = y <sub>t-3</sub> )	lag 1 (y <sub>4</sub> = y <sub>t-4</sub> )	lag 5 (y <sub>5</sub> = y <sub>t-5</sub> )	y <sub>t-m</sub>		y1-m	y2-m	y3-m				
1	1230	1345	1382	1416	1593	1802	-1010.2		-895.2	-858.2	-858.2				
2	1345	1382	1416	1593	1802	1817	-895.2		-858.2	-824.2	-824.2				
3	1382	1416	1593	1802	1817	1995	-858.2		-824.2	-647.2	-647.2				
4	1416	1593	1802	1817	1995	2212	-824.2		-647.2	-438.2	-438.2				
5	1593	1802	1817	1995	2212	2607	-647.2		-438.2	-423.2	-423.2				
6	1802	1817	1995	2212	2607	2698	-438.2		-423.2	-245.2	-245.2				
7	1817	1995	2212	2607	2698	2703	-423.2		-245.2	-28.2	-28.2				
8	1995	2212	2607	2698	2703	2738	-245.2		-28.2	366.8	366.8				
9	2212	2607	2698	2703	2738	2810	-28.2		366.8	457.8	457.8				
10	2607	2698	2703	2738	2810	2632	366.8		457.8	462.8	462.8				
11	2698	2703	2738	2810	2632	2548	457.8		462.8	497.8	497.8				
12	2703	2738	2810	2632	2548	2881	462.8		497.8	569.8	569.8				
13	2738	2810	2632	2548	2881	2725	497.8		569.8	391.8	391.8				
14	2810	2632	2548	2881	2725	2911	569.8		391.8	307.8	307.8				
15	2632	2548	2881	2725	2911	2759	391.8		307.8	640.8	640.8				
16	2548	2881	2725	2911	2759		307.8		640.8	484.8	484.8				



Excel File Edit View Insert Format Tools Data Window Help Prof.Gangabralah B Andanalal PhD Sat 5 Mar 8:14 PM

Time series data

Home Insert Draw Page Layout Formulas Data Review View

Cut Copy Format Calibri (Body) 22 A A = Wrap Text Number Conditional Formatting Auto-sum Paste Merge & Centre % Insert Delete Format Clear Sort & Filter

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

fx 1230

3	1382	1416	1593	1802	1817	1995	-858.2	-824.2	-647.2
4	1416	1593	1802	1817	1995	2212	-824.2	-647.2	-438.2
5	1593	1802	1817	1995	2212	2607	-647.2	-438.2	-423.2
6	1802	1817	1995	2212	2607	2698	-438.2	-423.2	-245.2
7	1817	1995	2212	2607	2698	2703	-423.2	-245.2	-28.2
8	1995	2212	2607	2698	2703	2738	-245.2	-28.2	366.8
9	2212	2607	2698	2703	2738	2810	-28.2	366.8	457.8
10	2607	2698	2703	2738	2810	2632	366.8	457.8	462.8
11	2698	2703	2738	2810	2632	2548	457.8	462.8	497.8
12	2703	2738	2810	2632	2548	2881	462.8	497.8	569.8
13	2738	2810	2632	2548	2881	2725	497.8	569.8	391.8
14	2810	2632	2548	2881	2725	2911	569.8	391.8	307.8
15	2632	2548	2881	2725	2911	2759	391.8	307.8	640.8
16	2548	2881	2725	2911	2759		307.8	640.8	484.8
17	2881	2725	2911	2759			640.8	484.8	670.8
18	2725	2911	2759				484.8	670.8	518.8
19	2911	2759					670.8	518.8	
20	2759								

Sheet1 Sheet2 208 x 10 Virtual Classroom

Arranging CHARTS Count: 70 Sum: 20300

100%



Time series data

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Yes No

Year (t)	$y_t$	lag 1 ( $y_1 = y_{t-1}$ )	lag 2 ( $y_2 = y_{t-2}$ )	lag3 ( $y_3 = y_{t-3}$ )	lag 1 ( $y_4 = y_{t-4}$ )	lag 5 ( $y_5 = y_{t-5}$ )	$y_{t-m}$	$y_{1-m}$	$y_{2-m}$	$y_{3-m}$
1	1230	1845	1382	1416	1593	1602	-1010.2	-895.2	-852.2	
2	1809	1982	1415	2393	2602	2817	-893.2	-838.2	-824.2	
3	1982	1415	1593	2602	2617	1993	-838.2	-824.2	-647.2	
4	1315	1593	2602	2617	2993	2712	-826.2	-647.2	-436.2	
5	2393	2602	1817	1995	2212	2602	-642.2	-438.2	-423.2	
6	1602	1817	1995	2212	2607	2698	-456.2	-423.2	-245.2	
7	1817	1993	2212	2607	2698	2703	-423.2	-245.2	-28.2	
8	2993	2212	2607	2698	2703	2738	-243.2	-28.2	566.8	
9	2212	2607	2698	2703	2738	2810	28.2	566.8	497.8	
10	2807	2638	2703	2738	2820	2632	366.8	497.8	407.8	
11	2632	2703	2738	2810	2632	2549	457.8	467.8	497.8	
12	2909	2738	2810	2632	2548	2881	462.8	497.8	569.8	
13	2738	2810	2632	2548	2881	3725	497.8	569.8	591.8	
14	2810	2511	2632	2549	2881	2725	2911	589.8	891.8	307.8
15	2511	2632	2549	2881	2725	2911	7759	491.8	307.8	641.9

Average: 2340.2 Count: 26 Sum: 63804

