SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

COURSE NAME : 19CS407-DATA ANALYTICS WITH R

II YEAR /IV SEMESTER

Unit II – Statistics and Prescriptive Analytics

Topic : TIME SERIES ANALYSIS

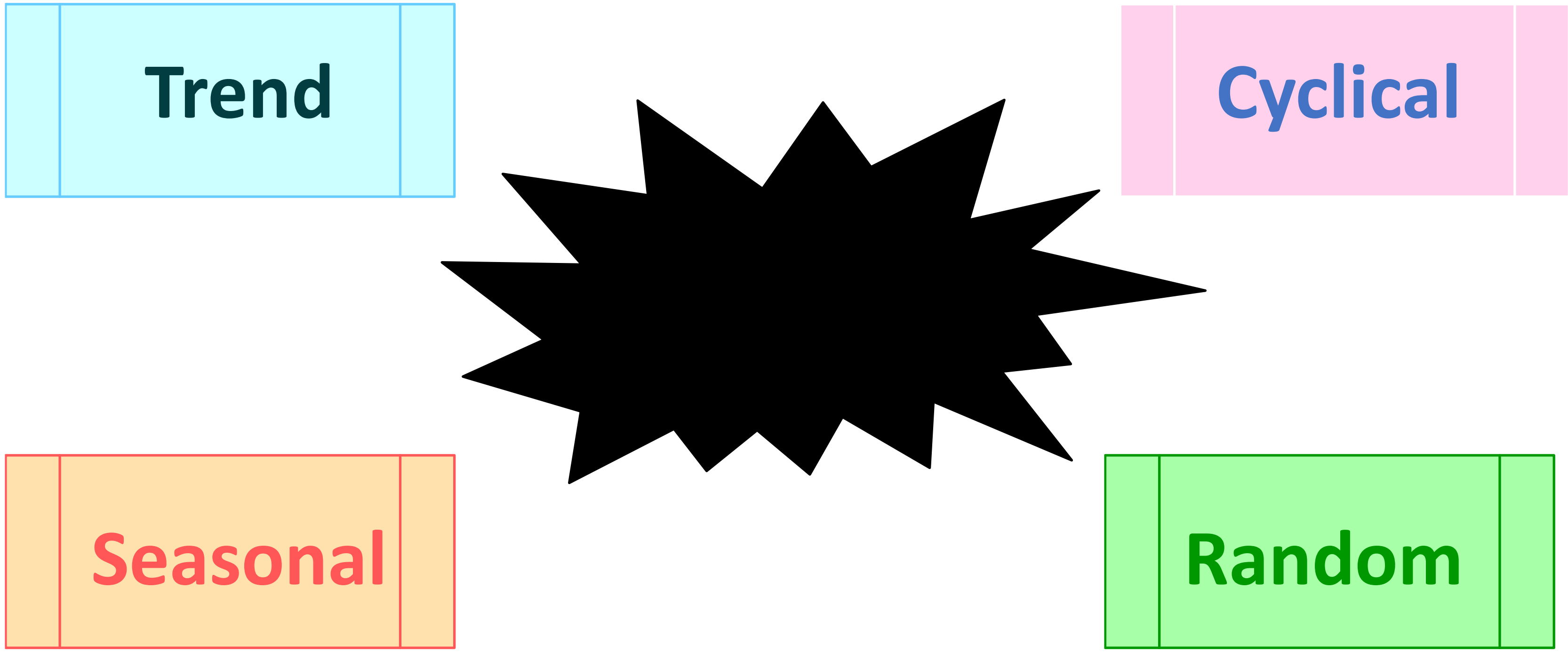
- Component Factors of the Time-Series Model
- Smoothing of Data Series
  - Moving Averages
  - Exponential Smoothing
- Least Square Trend Fitting and Forecasting
  - Linear, Quadratic and Exponential Models
- Autoregressive Models
- Choosing Appropriate Models
- Monthly or Quarterly Data

What Is Time-Series

- A Quantitative Forecasting Method to Predict Future Values
- Numerical Data Obtained at Regular Time Intervals
- Projections Based on Past and Present Observations
- Example:

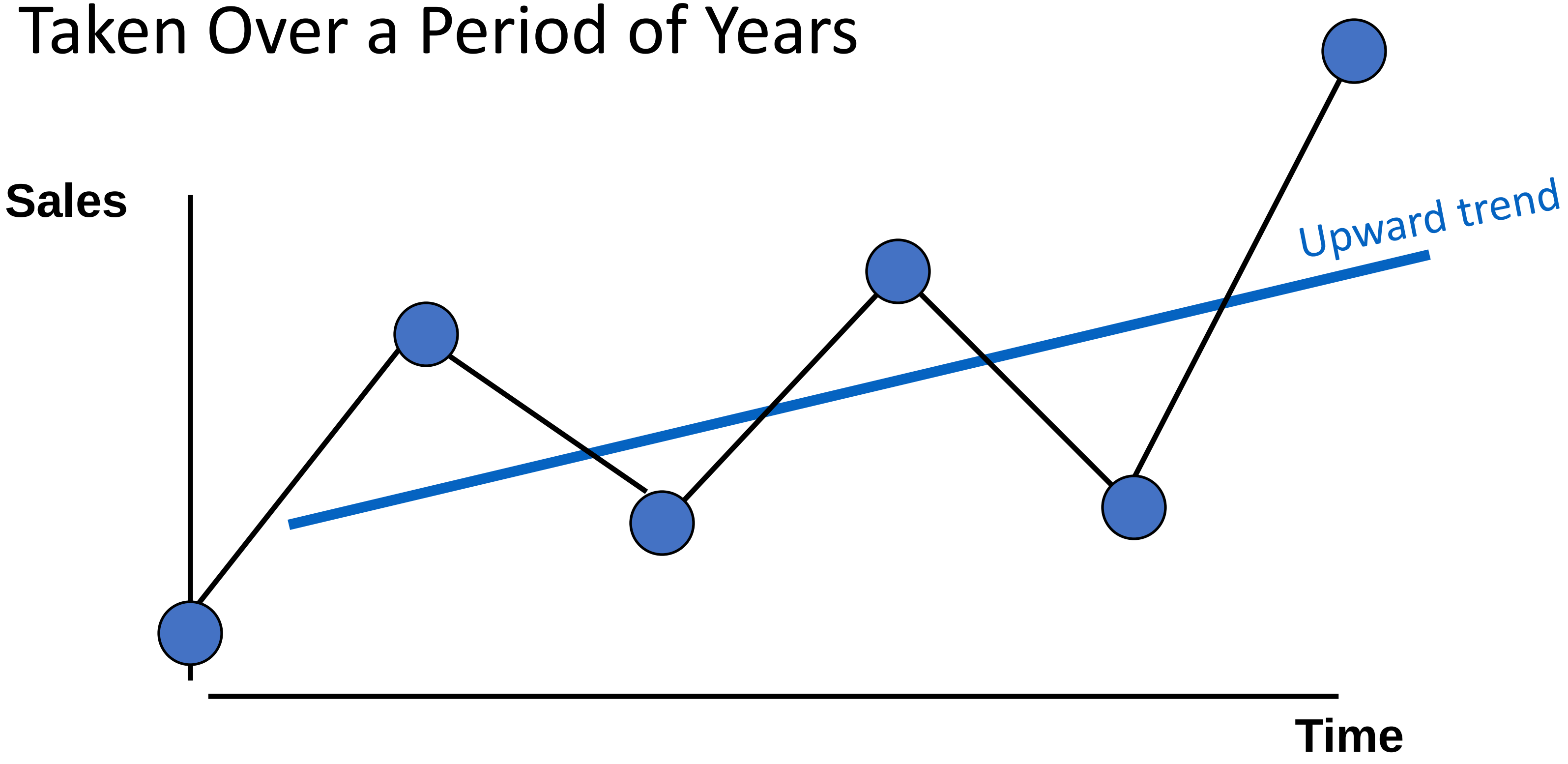
Year:	1994	1995	1996	1997	1998
Sales:	75.3	74.2	78.5	79.7	80.2

Time-Series Components



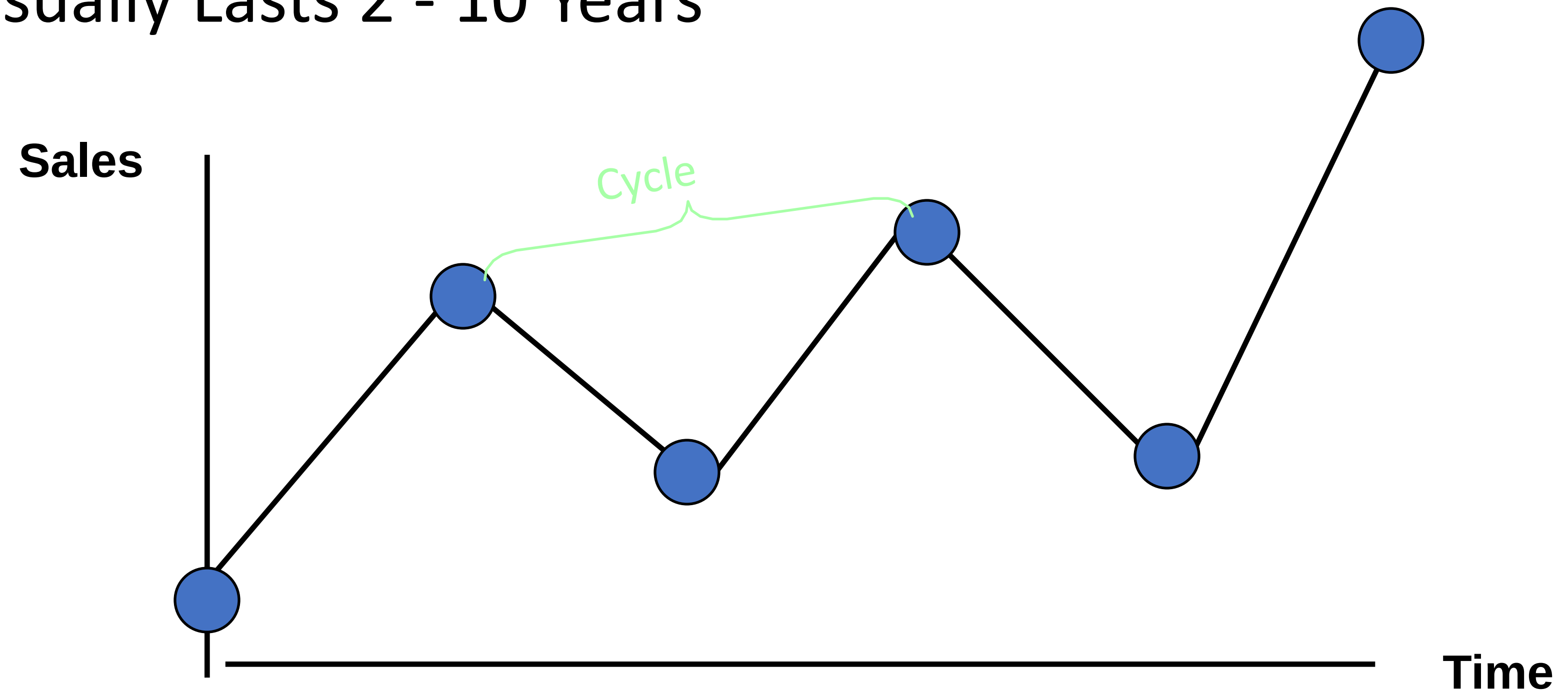
Trend Component

- Overall Upward or Downward Movement
- Data Taken Over a Period of Years



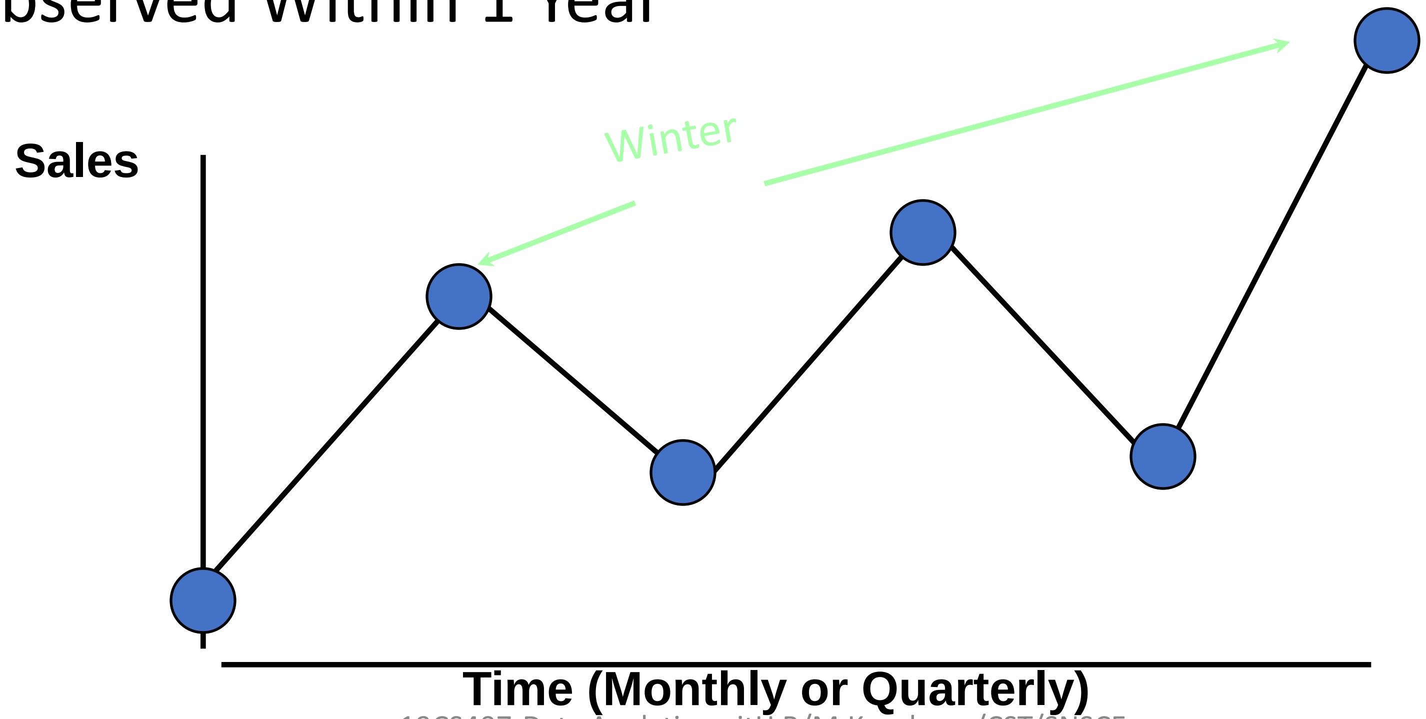
Cyclical Component

- Upward or Downward Swings
- May Vary in Length
- Usually Lasts 2 - 10 Years



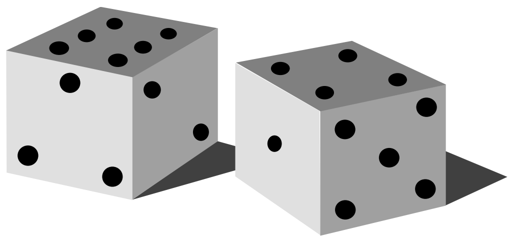
Seasonal Component

- Upward or Downward Swings
- Regular Patterns
- Observed Within 1 Year



Random or Irregular Component

- Erratic, Nonsystematic, Random, ‘Residual’ Fluctuations
- Due to Random Variations of
  - Nature
  - Accidents
- Short Duration and Non-repeating



Multiplicative Time-Series Model

- Used Primarily for Forecasting
- Observed Value in Time Series is the product of Components
- For Annual Data:

$$Y_i = T_i \times C_i \times I_i$$

- For Quarterly or Monthly Data:

$$Y_i = T_i \times S_i \times C_i \times I_i$$

T_i = Trend

C_i = Cyclical

I_i = Irregular

S_i = Seasonal

Moving Averages

- Used for Smoothing
- Series of Arithmetic Means Over Time
- Result Dependent Upon Choice of *L*, Length of Period for Computing Means
- For Annual Time-Series, *L* Should be Odd
- Example: 3-year Moving Average
  - First Average:

$$MA(3) = \frac{Y_1 + Y_2 + Y_3}{3}$$

  - Second Average:

$$MA(3) = \frac{Y_2 + Y_3 + Y_4}{3}$$

Moving Average Example

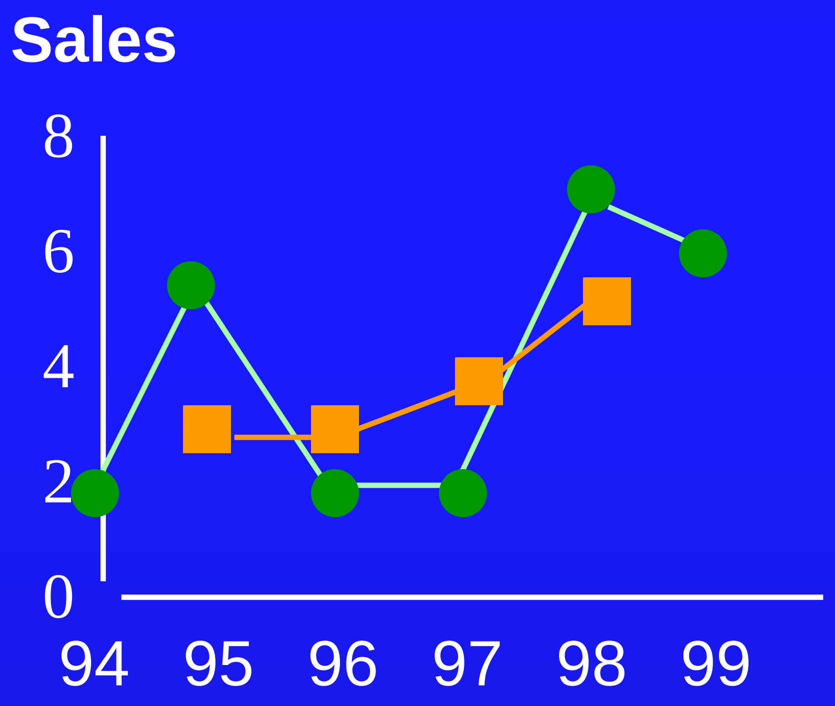
John is a building contractor with a record of a total of 24 single family homes constructed over a 6 year period.

Provide John with a Moving Average Graph.

Year	Units	Moving Ave
1994	2	NA
1995	5	3
1996	2	3
1997	2	3.67
1998	7	5
1999	6	NA

Moving Average Example Solution

Year	Response ●	Moving Ave ■
1994	2	NA
1995	5	3
1996	2	3
1997	2	3.67
1998	7	5
1999	6	NA



© 1999 Prentice-Hall, Inc.

Exponential Smoothing

- Weighted Moving Average
  - Weights Decline Exponentially
  - Most Recent Observation Weighted Most
- Used for Smoothing and Short Term Forecasting
- Weights Are:
  - Subjectively Chosen
  - Ranges from 0 to 1
    - Close to 0 for Smoothing
    - Close to 1 for Forecasting

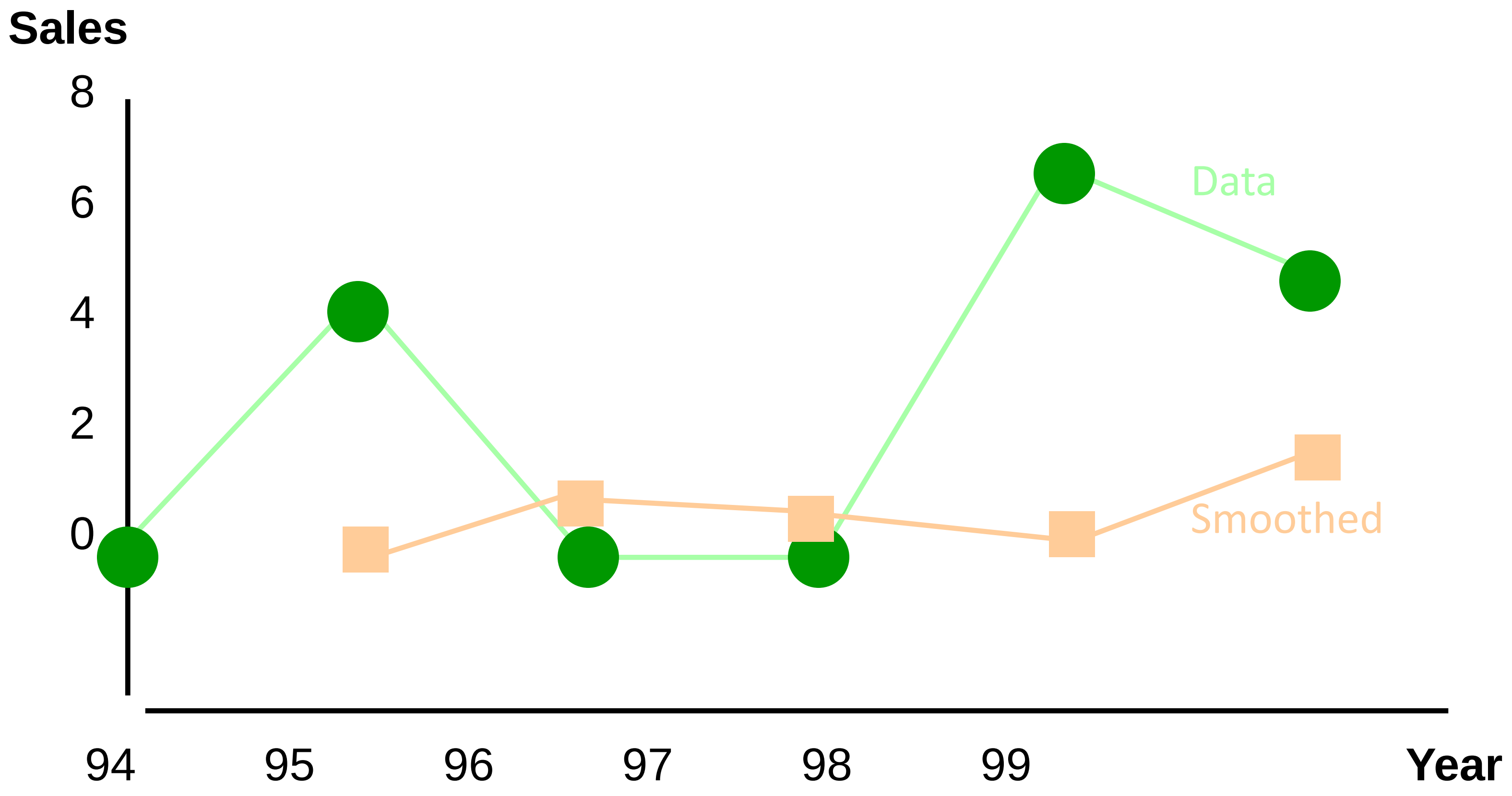
Exponential Weight: Example

$$E_i = WY_i + (1-W)E_{i-1}$$

Year	Response	Smoothing Value (W = .2)	Forecast
1994	2	2	NA
1995	5	(.2)(5) + (.8)(2) = 2.6	**2**
1996	2	(.2)(2) + (.8)(2.6) = 2.48	**2.6**
1997	2	(.2)(2) + (.8)(2.48) = 2.384	**2.48**
1998	7	(.2)(7) + (.8)(2.384) = 3.307	**2.384**
1999	6	(.2)(6) + (.8)(3.307) = 3.846	**3.307**

© 1999 Prentice-Hall, Inc.

Exponential Weight: Example Graph



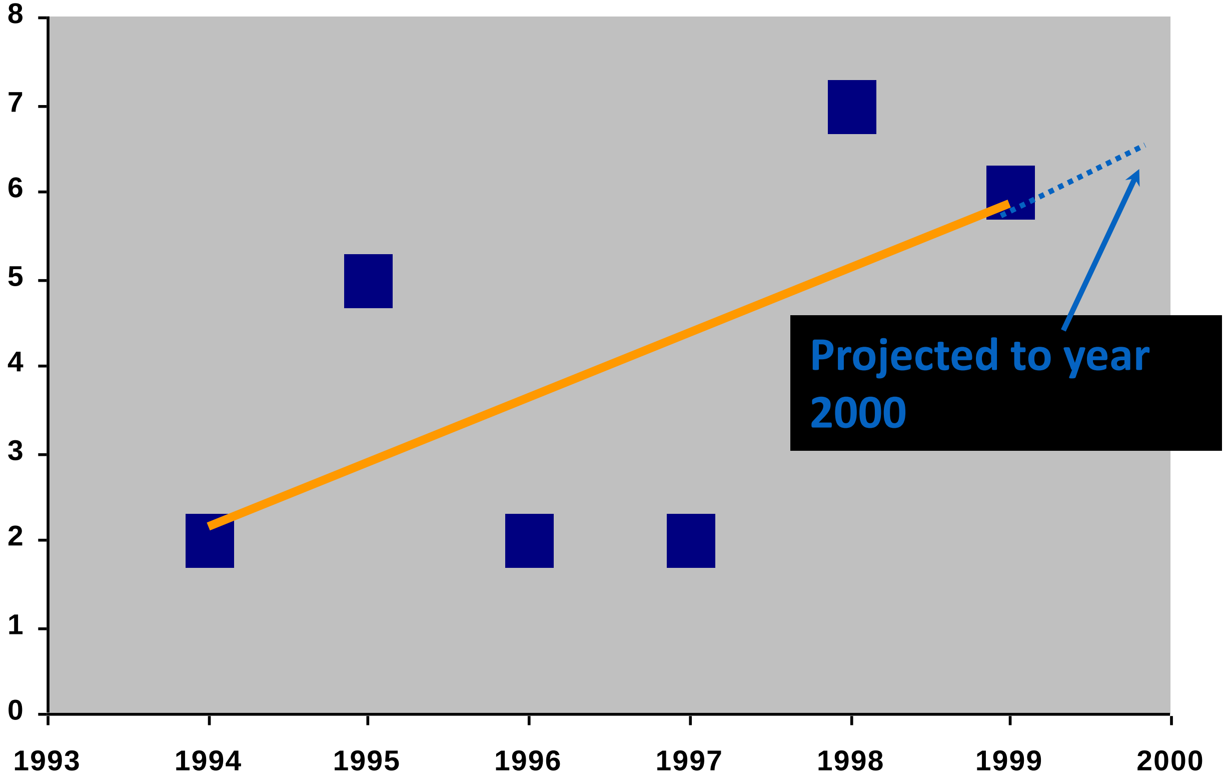
The Linear Trend Model

Year	Coded	Sales
94	0	2
95	1	5
96	2	2
97	3	2
98	4	7
99	5	6

$$\hat{Y}_i = b_0 + b_1 X_i = 2.143 + .743 X_i$$

Excel Output

	Coefficients
Intercept	2.14285714
X Variable 1	0.74285714



The Quadratic Trend Model

Year	Coded	Sales
94	0	2
95	1	5
96	2	2
97	3	2
98	4	7
99	5	6

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 X_i^2$$

	Coefficients
Intercept	2.85714286
X Variable 1	-0.3285714
X Variable 2	0.21428571

Excel Output

$$\hat{Y}_i = 2.857 - 0.33 X_i + .214 X_i^2$$

Autogregressive Modeling

- Used for Forecasting
- Takes Advantage of Autocorrelation
  - 1st order - correlation between consecutive values
  - 2nd order - correlation between values 2 periods apart
- Autoregressive Model for *pth* order:

Random Error

$$Y_i = A_0 + A_1 Y_{i-1} + A_2 Y_{i-2} + \bullet\bullet\bullet + A_p Y_{i-p} + \delta_i$$

Autoregressive Model: Example

The Office Concept Corp. has acquired a number of office units (in thousands of square feet) over the last 8 years. Develop the 2nd order Autoregressive models.

Year	Units
92	4
93	3
94	2
95	3
96	2
97	2
98	4
99	6

Autoregressive Model: Example Solution

- Develop the 2nd order table
- Use Excel to run a regression model

Year	Y_i	Y_{i-1}	Y_{i-2}
92	4	---	---
93	3	4	---
94	2	3	4
95	3	2	3
96	2	3	2
97	2	2	3
98	4	2	2
99	6	4	2

Excel Output

	Coefficients
Intercept	3.5
X Variable 1	0.8125
X Variable 2	-0.9375

$$Y_i = 3.5 + .8125 Y_{i-1} - .9375 Y_{i-2}$$

Autoregressive Model Example: Forecasting

Use the 2nd order model to forecast number of units for 2000:

$$Y_i = 3.5 + .8125Y_{i-1} - .9375Y_{i-2}$$

$$\begin{aligned} Y_{2000} &= 3.5 + .8125Y_{1999} - .9375Y_{1998} \\ &= 3.5 + .8125 \times 6 - .9375 \times 4 \\ &= 4.625 \end{aligned}$$

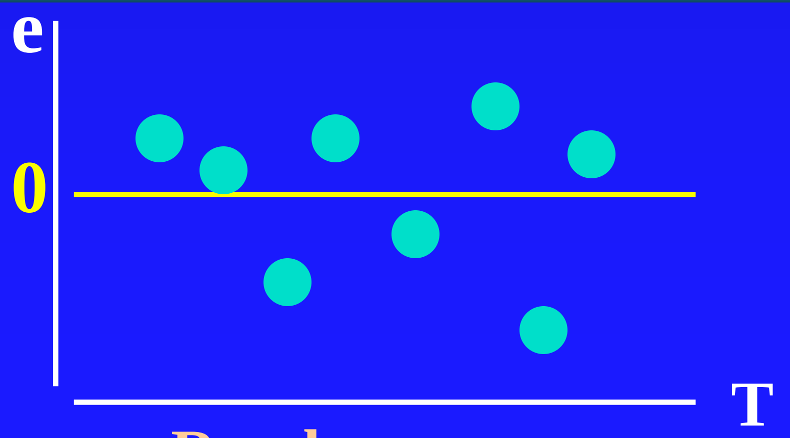
Autoregressive Modeling Steps

- 1. Choose *p*: Note that df = n - 2*p* - 1
- 2. Form a series of "lag predictor" variables
- Y_{i-1} , Y_{i-2} , … Y_{i-p}
- 3. Use Excel to run regression model using all *p* variables
- 4. Test significance of A_p
  - If null hypothesis rejected, this model is selected
  - If null hypothesis not rejected, decrease *p* by 1 and repeat

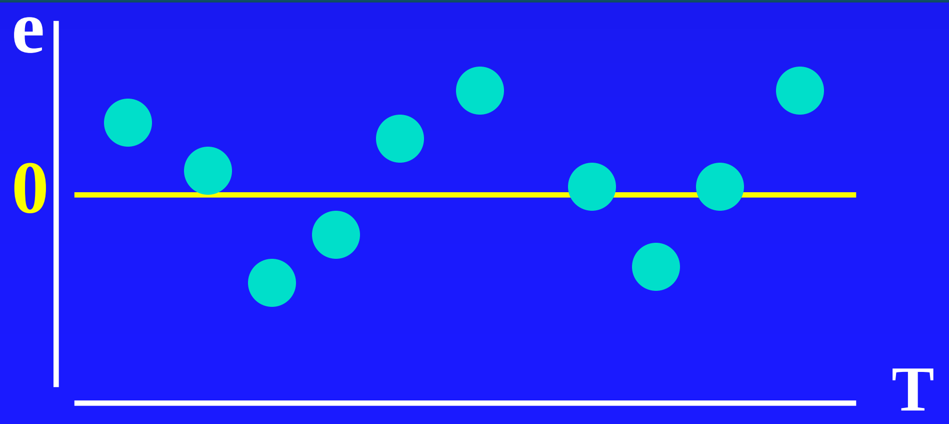
Selecting A Forecasting Model

- Perform A Residual Analysis
  - Look for pattern or direction
- Measure Sum Square Errors - SSE (residual errors)
- Measure Residual Errors Using MAD
- Use Simplest Model
  - Principle of Parsimony

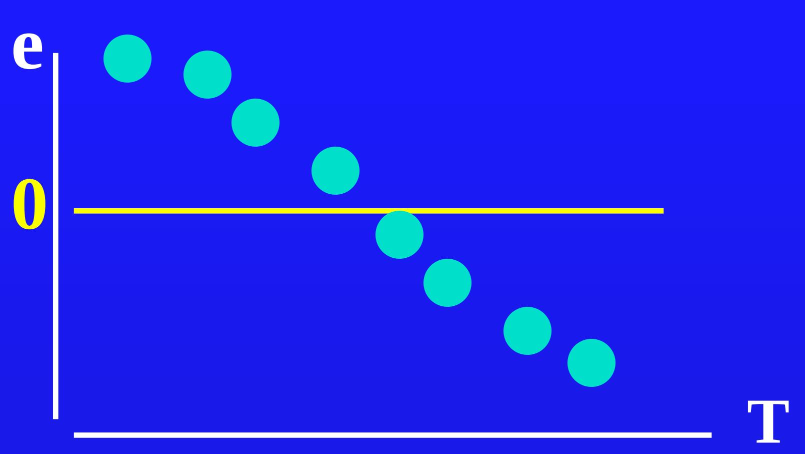
Residual Analysis



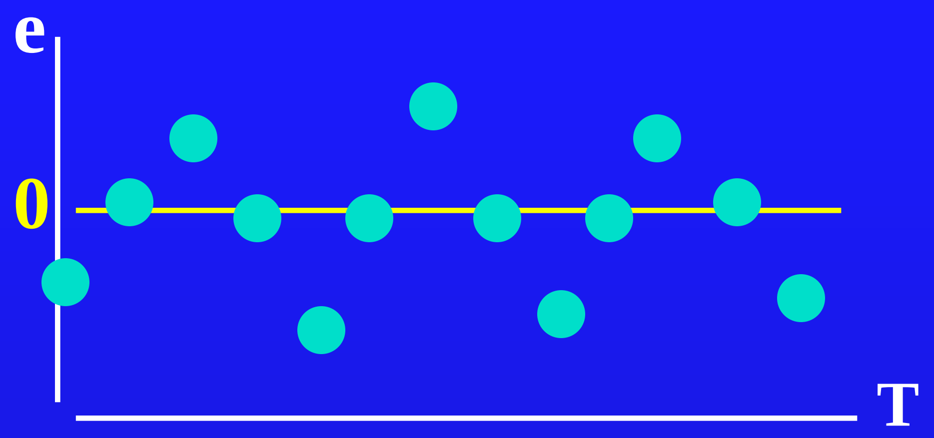
Random errors



Cyclical effects not accounted for



Trend not accounted for



Seasonal effects not accounted for

© 1999 Prentice-Hall, Inc.

Measuring Errors

- Sum Square Error (SSE)

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

- Mean Absolute Deviation (MAD)

$$MAD = \frac{\sum_{i=1}^{n} \left| Y_i - \hat{Y}_i \right|}{n}$$